Final_Project Report

Ananya, Rosie, Sameer

2025-04-04

```
#Load libraries
library(tidyverse)
library(forecast)
#library(tseries)
library(sarima)
library(lubridate)
library(Kendall)
library(outliers)
library(foreign)
library(here)
library(cowplot)
library(kableExtra)
library(smooth)
library(tseries)
```

Introduction

Why are we interested in Tidal Power?

Relevance to our fields of studies and interests:

Our group is made up with students from Coastal Marine Systems and Energy concentrations under the MEM program, and we decided that this topic covered the fields of interests for all of our group members.

Great Potential of Tidal Energy:

Clean & Sustainable Energy Source:
- Tidal power is a renewable energy that produces zero greenhouse gas emissions. - Unlike traditional hydro power, tidal energy does not require dams or large reservoirs, making it more environmentally friendly.

High Energy Density & Efficiency:
- Wind speeds shift, and sunlight fades, so such systems require storage to sustain their service areas' electrical demand.
- By contrast, ocean currents are relatively constant, and some amount of motion still occurs at all times.
- So, a tidal energy farm can provide continuous power without the need for expensive battery or fuel cell storage.
- Tidal energy has a higher power capacity than many other renewable, helping bridge the gap in clean energy demand.

Reliable & Predictable:
- Unlike intermittent sources like wind and solar, tidal currents are consistent and predictable, providing continuous power without heavy reliance on storage.
- This reliability helps balance energy grids, complementing other renewables during low-generation periods.

Growing U.S. Interest & Investment:
- No commercially available sites right now in the US.
- There are Early projects in Maine and New York have successfully tested tidal turbines.

Project Overview:

In this project, we are aiming to compare the Tidal Power Potential across three sites in different latitudes and coastal areas with different geographic and climatic conditions in the US. The tidal power potential is assessed by:
- The Wave Energy Flux (Watts per square meter) = Wave Power Density (Joules per cubic meter) * Wave Energy Period (seconds).
- Then in the Analysis below, we are going to compare the seasonal and long term trends of the time series of the Wave Energy Flux that we calculated.
- Then we will run models of prediction based on training and testing datasets that we extracted in the original time series data.
- Then we compare the model forecasting results to see which site has better model prediction results and higher/ more stable values (Wave Energy Flux).

Data selection

We first picked 3 sites/ locations/ areas tend to be some of the locations referred to for its tidal power potential in the general Internet searches. These three sites are:
- Gulf of Maine (43.68° N, 69.77 ° W)
- ~ 3 km distance into ocean near Miami, Florida (25.74 ° N, 80.11 ° W)
- Cook Inlet, Alaska (59.37° N, 152.64° W)

With the lat-long of these sites, we extracted raw data of 1979-2010 Wave Energy Density and Energy period data in NREL marine energy atlas: http://nrel.gov/marine-energy-atlas/data-viewer

These areas are distinct by latitudes and longitudes, as well as different coasts and climatic environments.

Data Wrangling & Initial Analysis

- In general, we first calculated the Wave Energy Flux data based on the raw data we extracted.
- Then the Wave Energy Flux datasets for all 3 sites' were aggregated to daily and monthly view (original series).
- Then we split the current full dataset using 80/20 rule into training and testing datasets.
- Then, we conducted the original time series analysis.
- Then we ran various models to compare predicting/ forecasting performance on the Maine dataset.
- Then we used the top 3 performance models (Arima + Fourier k-4, TBATs. and STL+ETS) for the other 2 sites Wave Energy Flux.
- Compare Wave Energy Flux Results of the 3 sites and determine which site has more potential.

```
#Import datasets
here()
```

```
## [1] "/Users/sameerswarup/Documents/Duke MEM Classes/Spring 2025/Time Series Analysis for Energy/RStu
```

```
getwd()
```

```
## [1] "/Users/sameerswarup/Documents/Duke MEM Classes/Spring 2025/Time Series Analysis for Energy/RStuc
```

```r
alaska_df <- read.csv("../Data/Alaska_combined_output.csv",
                      skip = 2,
                      header = TRUE)
florida_df <- read.csv("../Data/Florida_combined_output.csv",
                       skip = 2,
                       header = TRUE)
maine_df <- read.csv("../Data/Gulf of Maine_combined_output.csv",
                     skip = 2,
                     header = TRUE)
```

```r
#Create a column that contains daily mean energy period, wave power and wave height and compute mean wa
alaska_daily_df <- alaska_df %>%
              group_by(Year, Month, Day) %>%
              summarise(daily_mean_wave_energy = mean(Energy.Period, na.rm = TRUE),
                        daily_mean_wave_height = mean(Significant.Wave.Height, na.rm = TRUE),
                        daily_mean_wave_power = mean(Omni.Directional.Wave.Power, na.rm = TRUE))
```

```
## `summarise()` has grouped output by 'Year', 'Month'. You can override using the
## `.groups` argument.
```

```r
florida_daily_df <- florida_df %>%
              group_by(Year, Month, Day) %>%
              summarise(daily_mean_wave_energy = mean(Energy.Period, na.rm = TRUE),
                        daily_mean_wave_height = mean(Significant.Wave.Height, na.rm = TRUE),
                        daily_mean_wave_power = mean(Omni.Directional.Wave.Power, na.rm = TRUE))
```

```
## `summarise()` has grouped output by 'Year', 'Month'. You can override using the
## `.groups` argument.
```

```r
maine_daily_df <- maine_df %>%
              group_by(Year, Month, Day) %>%
              summarise(daily_mean_wave_energy = mean(Energy.Period, na.rm = TRUE),
                        daily_mean_wave_height = mean(Significant.Wave.Height, na.rm = TRUE),
                        daily_mean_wave_power = mean(Omni.Directional.Wave.Power, na.rm = TRUE))
```

```
## `summarise()` has grouped output by 'Year', 'Month'. You can override using the
## `.groups` argument.
```

```r
#Create a column that contains monthly mean energy period, wave power and wave height and compute mean w
alaska_monthly_df <- alaska_daily_df %>%
              group_by(Year, Month) %>%
              summarise(monthly_mean_wave_energy = mean(daily_mean_wave_energy, na.rm = TRUE),
                        monthly_mean_wave_height = mean(daily_mean_wave_height, na.rm = TRUE),
                        monthly_mean_wave_power = mean(daily_mean_wave_power, na.rm = TRUE))
```

```
## `summarise()` has grouped output by 'Year'. You can override using the
## `.groups` argument.
```

```r
florida_monthly_df <- florida_daily_df %>%
              group_by(Year, Month) %>%
              summarise(monthly_mean_wave_energy = mean(daily_mean_wave_energy, na.rm = TRUE),
                        monthly_mean_wave_height = mean(daily_mean_wave_height, na.rm = TRUE),
                        monthly_mean_wave_power = mean(daily_mean_wave_power, na.rm = TRUE))
```

```
## `summarise()` has grouped output by 'Year'. You can override using the
## `.groups` argument.
```

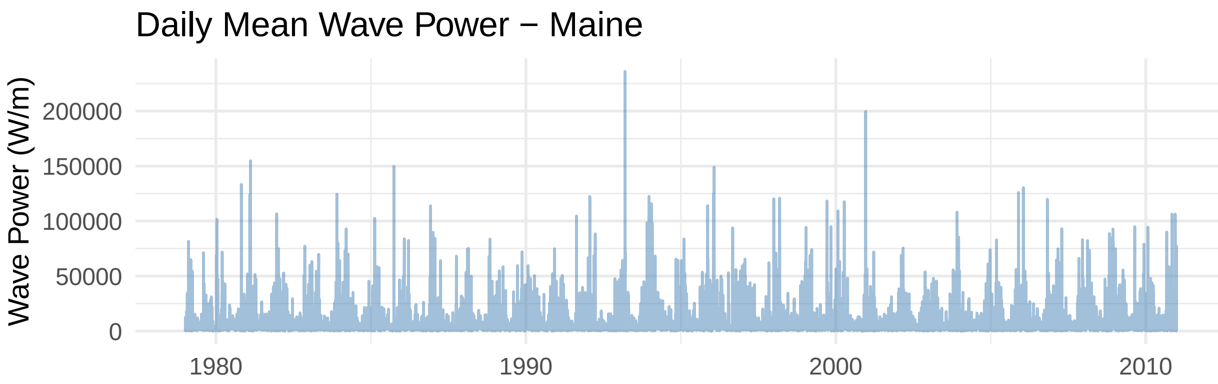
```r
maine_monthly_df <- maine_daily_df %>%
              group_by(Year, Month) %>%
              summarise(monthly_mean_wave_energy = mean(daily_mean_wave_energy, na.rm = TRUE),
                        monthly_mean_wave_height = mean(daily_mean_wave_height, na.rm = TRUE),
                        monthly_mean_wave_power = mean(daily_mean_wave_power, na.rm = TRUE))
```

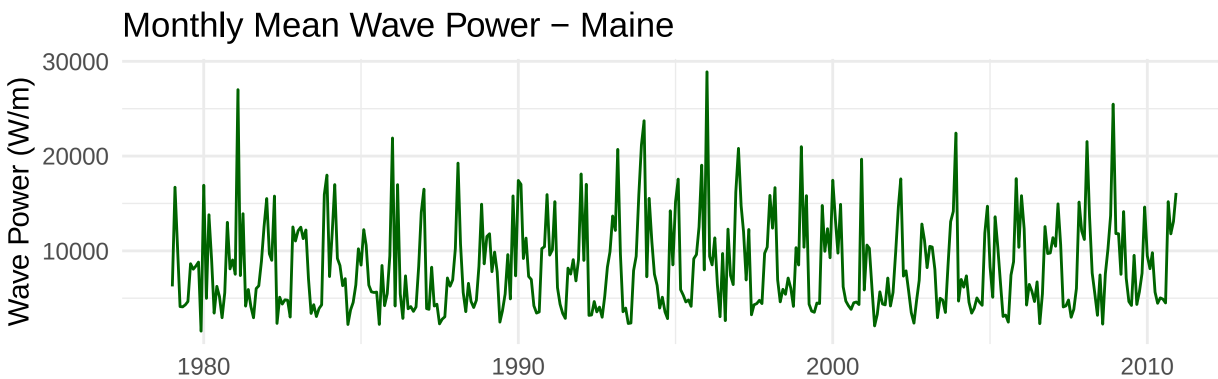
```
## `summarise()` has grouped output by 'Year'. You can override using the
## `.groups` argument.
```

Maine

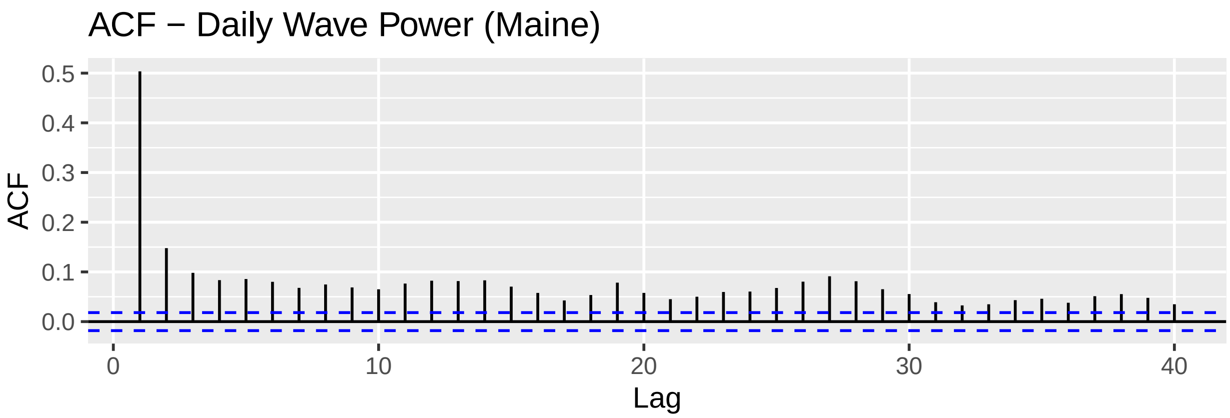
###We will use Maine to test our various time series models. Using accuracy metrics and plots we will choose the 3 best models and forecast mean monthly wave power for Alaska and Florida.

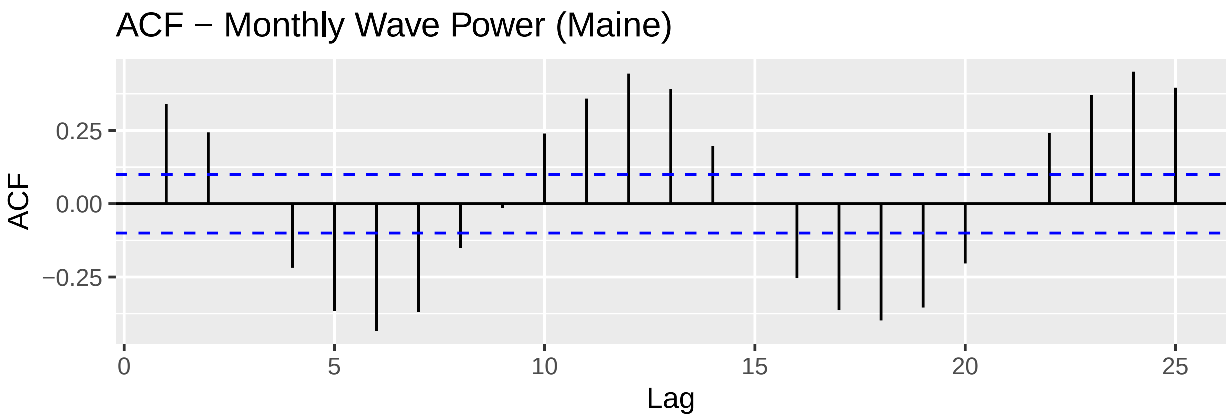
```r
maine_daily_df <- maine_daily_df %>%
  mutate(date = make_date(Year, Month, Day))

maine_monthly_df <- maine_monthly_df %>%
  mutate(date = make_date(Year, Month, 1))

#initial plots
maine_d_plot <- ggplot(maine_daily_df, aes(x = date, y = daily_mean_wave_power)) +
  geom_line(color = "steelblue", alpha = 0.5) +
  labs(title = "Daily Mean Wave Power - Maine", y = "Wave Power (W/m)", x = "") +
  theme_minimal()

maine_m_plot <- ggplot(maine_monthly_df, aes(x = date, y = monthly_mean_wave_power)) +
  geom_line(color = "darkgreen") +
  labs(title = "Monthly Mean Wave Power - Maine", y = "Wave Power (W/m)", x = "") +
  theme_minimal()

plot_grid(maine_d_plot, maine_m_plot, ncol = 1)
```






```
#acf plots
maine_d_acf <- ggAcf(maine_daily_df$daily_mean_wave_power) + ggtitle("ACF - Daily Wave Power (Maine)")
maine_m_acf <- ggAcf(maine_monthly_df$monthly_mean_wave_power) + ggtitle("ACF - Monthly Wave Power (Mai
plot_grid(maine_d_acf, maine_m_acf, ncol = 1)
```



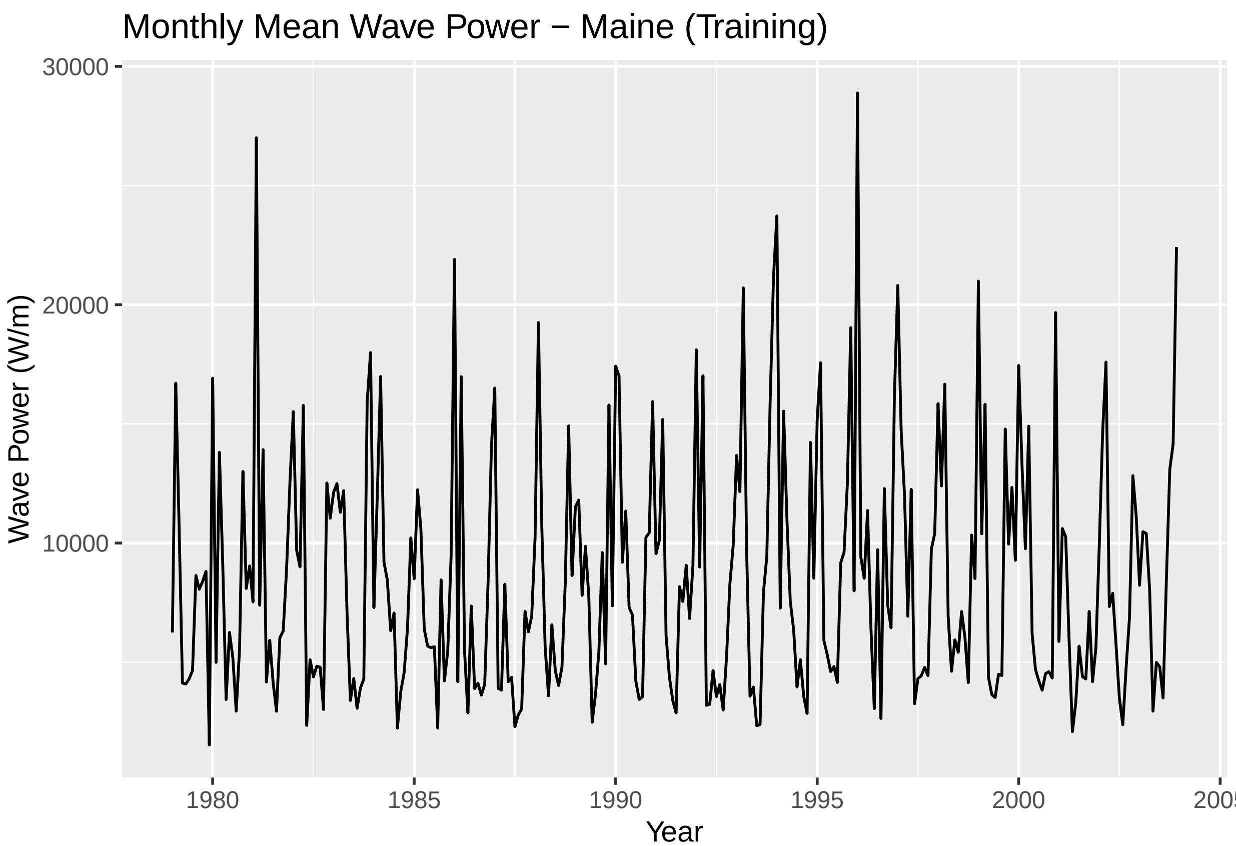


Observations: Daily mean wave power appears to be highly volatile and masked by noise, which might make it harder to model directly. On the other hand, monthly mean wave power shows a clear seasonal structure with regular peaks and dips, and consistent annual cycles across decades. Moreover, the ACF plot for the daily wave power drops sharply after lag 1, suggesting that past daily values don't carry much signal for future values. For the monthly wave power, the ACF plot shows strong seasonality with the wave pattern. The autocorrelation persists over time, which will be ideal for ARIMA/SARIMA and other seasonal models.

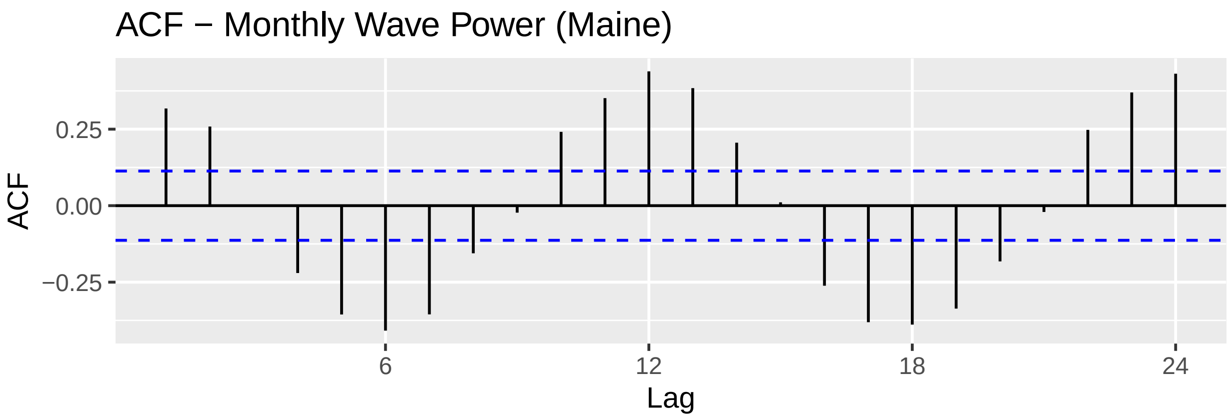
```
#Create training and test datasets using the 80-20 rule (for monthly data only)

#Training
alaska_monthly_train_df <- alaska_monthly_df %>%
                          filter(Year < 2004)
florida_monthly_train_df <- florida_monthly_df %>%
                          filter(Year < 2004)
maine_monthly_train_df <- maine_monthly_df %>%
                          filter(Year < 2004)

alaska_monthly_test_df <- alaska_monthly_df %>%
                          filter(Year > 2003)
florida_monthly_test_df <- florida_monthly_df %>%
                          filter(Year > 2003)
maine_monthly_test_df <- maine_monthly_df %>%
                          filter(Year > 2003)
```

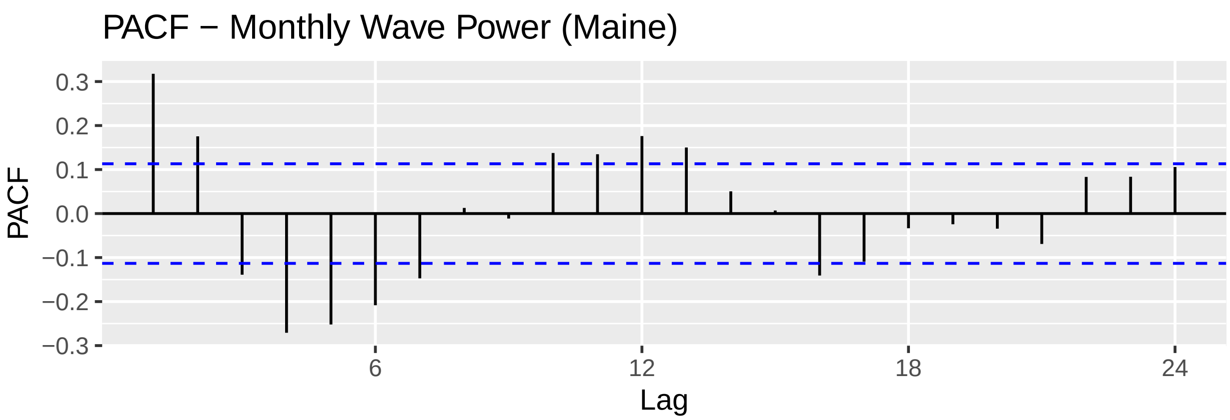
```
#converting to time series object
maine_ts_train <- ts(maine_monthly_train_df$monthly_mean_wave_power,
                     start = c(min(maine_monthly_train_df$Year), min(maine_monthly_train_df$Month)),
```

```r
                     frequency = 12)

maine_ts_test <- ts(maine_monthly_test_df$monthly_mean_wave_power,
                     start = c(min(maine_monthly_test_df$Year), min(maine_monthly_test_df$Month)),
                     frequency = 12)

maine_train_plot <- autoplot(maine_ts_train) +
  ggtitle("Monthly Mean Wave Power - Maine (Training)") +
  xlab("Year") + ylab("Wave Power (W/m)")
plot(maine_train_plot)
```

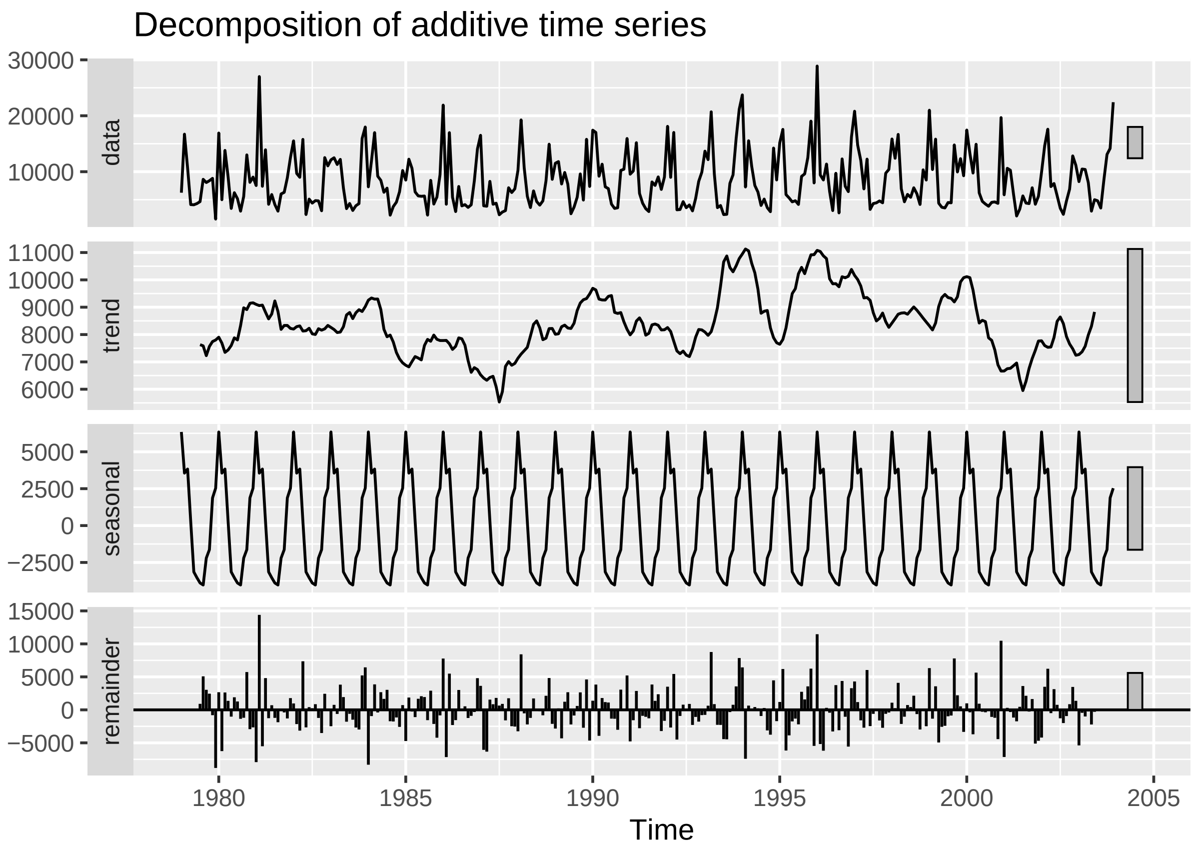



```r
#plotting acf and pacf
maine_train_acf <- ggAcf(maine_ts_train) + ggtitle("ACF - Monthly Wave Power (Maine)")
maine_train_pacf <- ggPacf(maine_ts_train) + ggtitle("PACF - Monthly Wave Power (Maine)")
plot_grid(maine_train_acf, maine_train_pacf, ncol = 1)
```

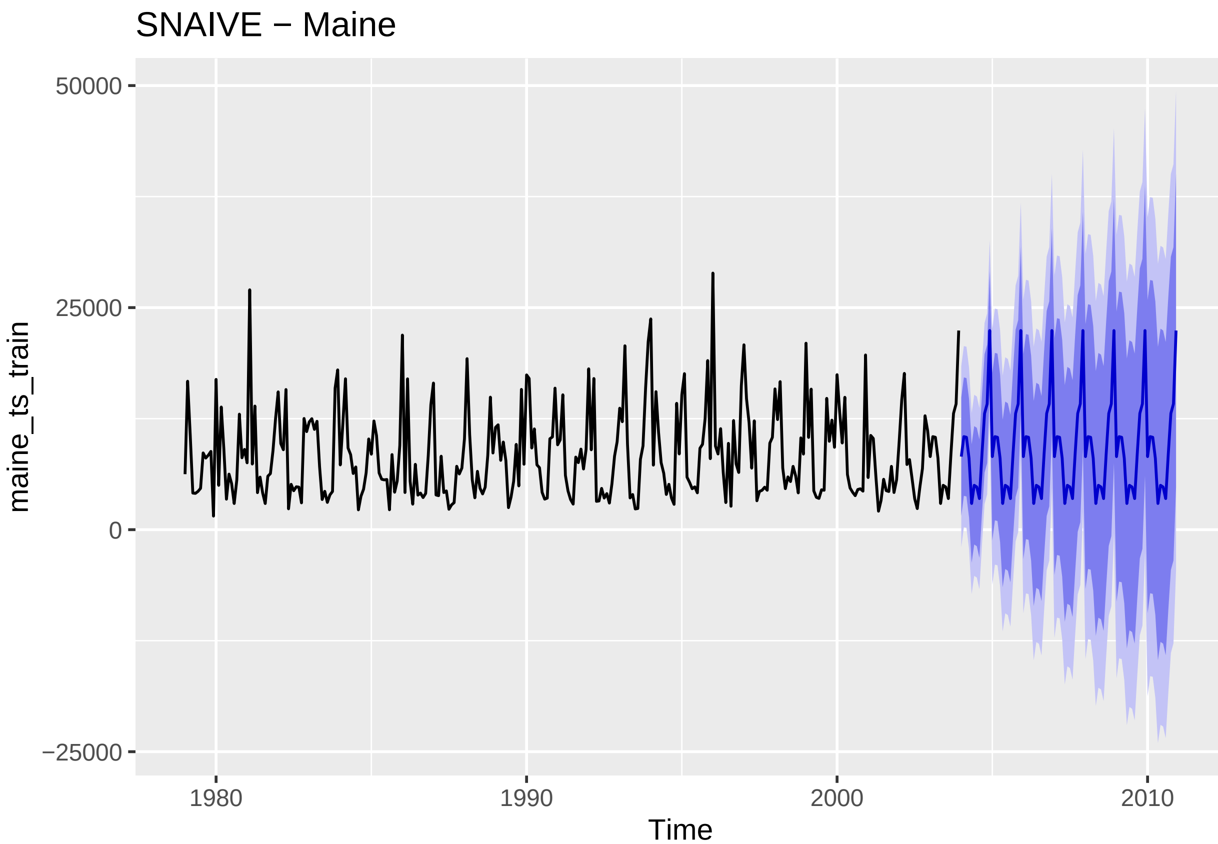





```
#decomposing the time series
maine_train_decomp <- decompose(maine_ts_train)
autoplot(maine_train_decomp)
```



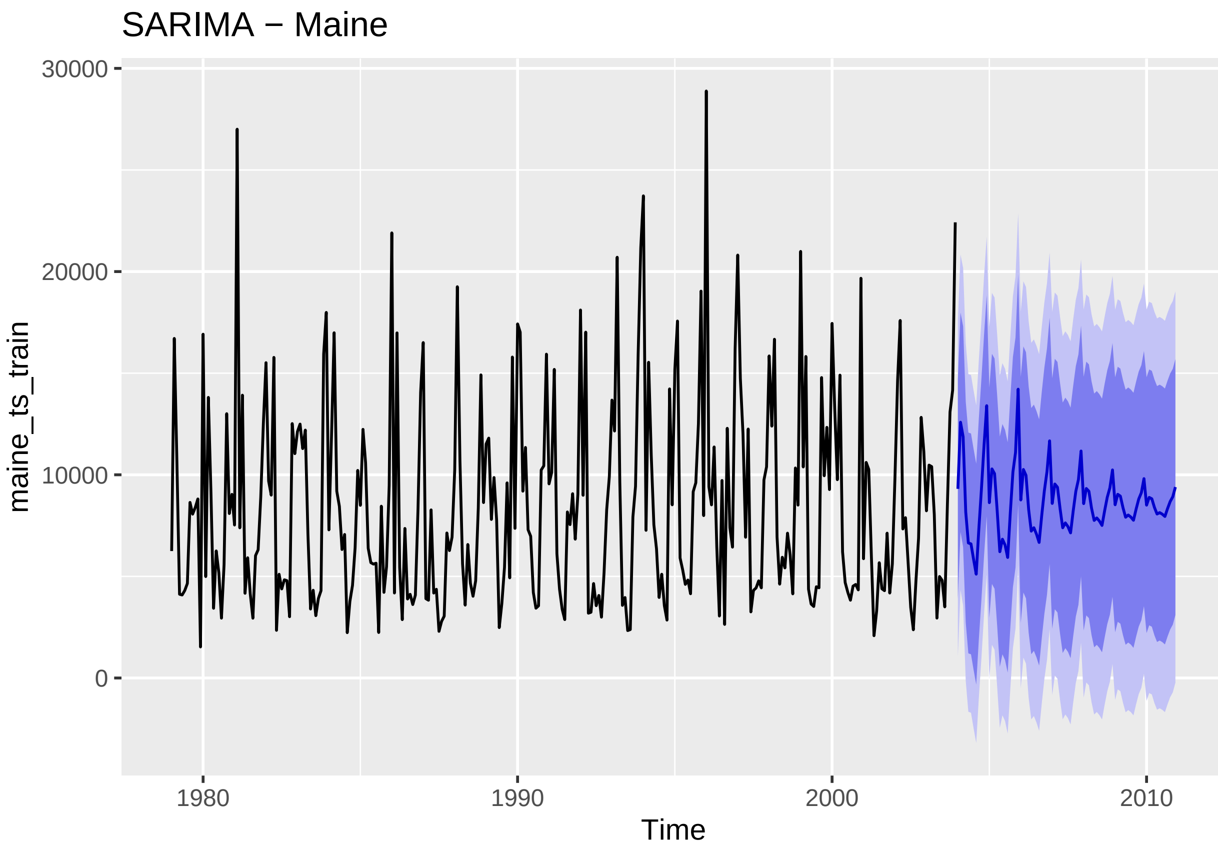

```r
#start of the forecasting

fore_h = length(maine_ts_test) #forecast horizon

#model 1: seasonal naive on original data (base model)
maine_snaive <- snaive(maine_ts_train, h = fore_h)
autoplot(maine_snaive) + ggtitle("SNAIVE - Maine")
```



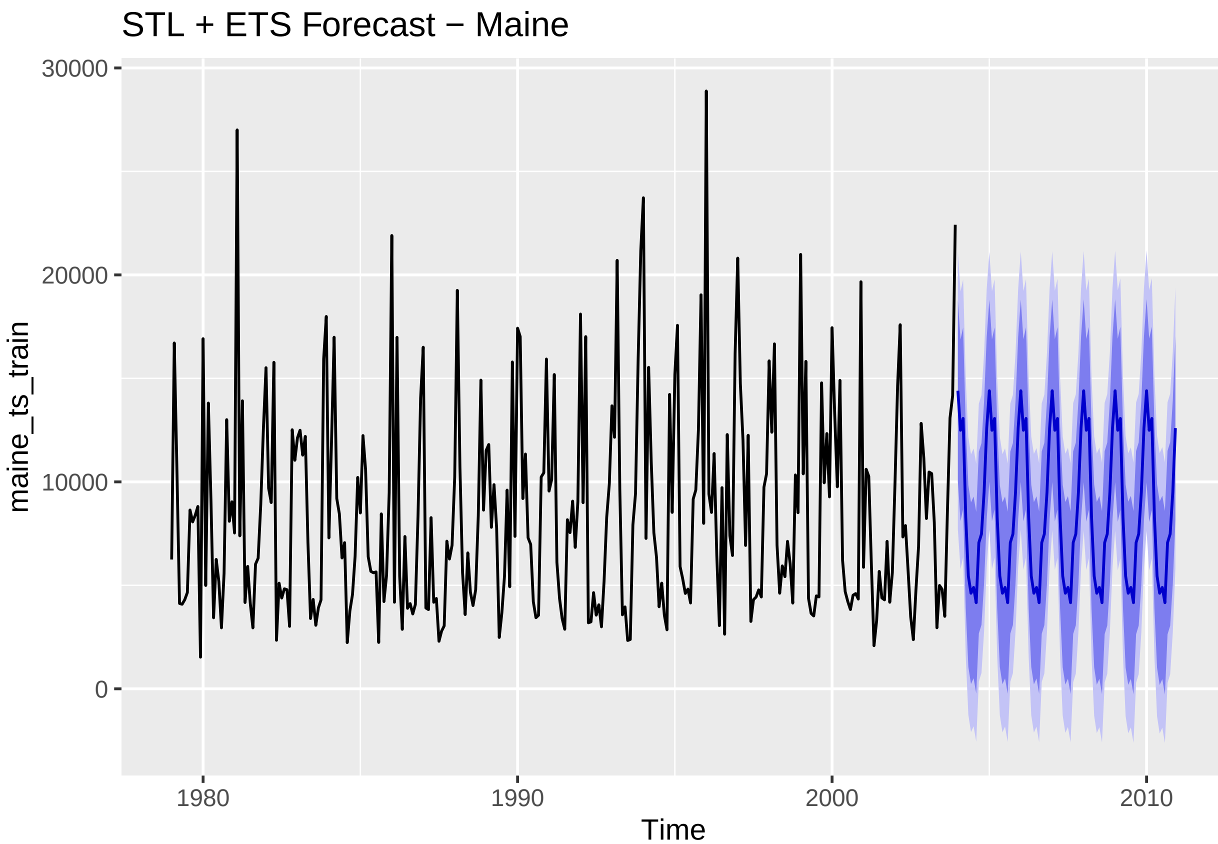

```
#model 2: SARIMA on original data
maine_sarima_fit <- auto.arima(maine_ts_train)
print(maine_sarima_fit)
```

```
## Series: maine_ts_train
## ARIMA(2,0,0)(2,0,0)[12] with non-zero mean
##
## Coefficients:
##           ar1     ar2    sar1    sar2       mean
##       -0.0136  0.1167  0.2955  0.3082  8431.9324
## s.e.   0.0661  0.0606  0.0573  0.0589   622.8468
##
## sigma^2 = 17720534:  log likelihood = -2929.12
## AIC=5870.23   AICc=5870.52   BIC=5892.45
```

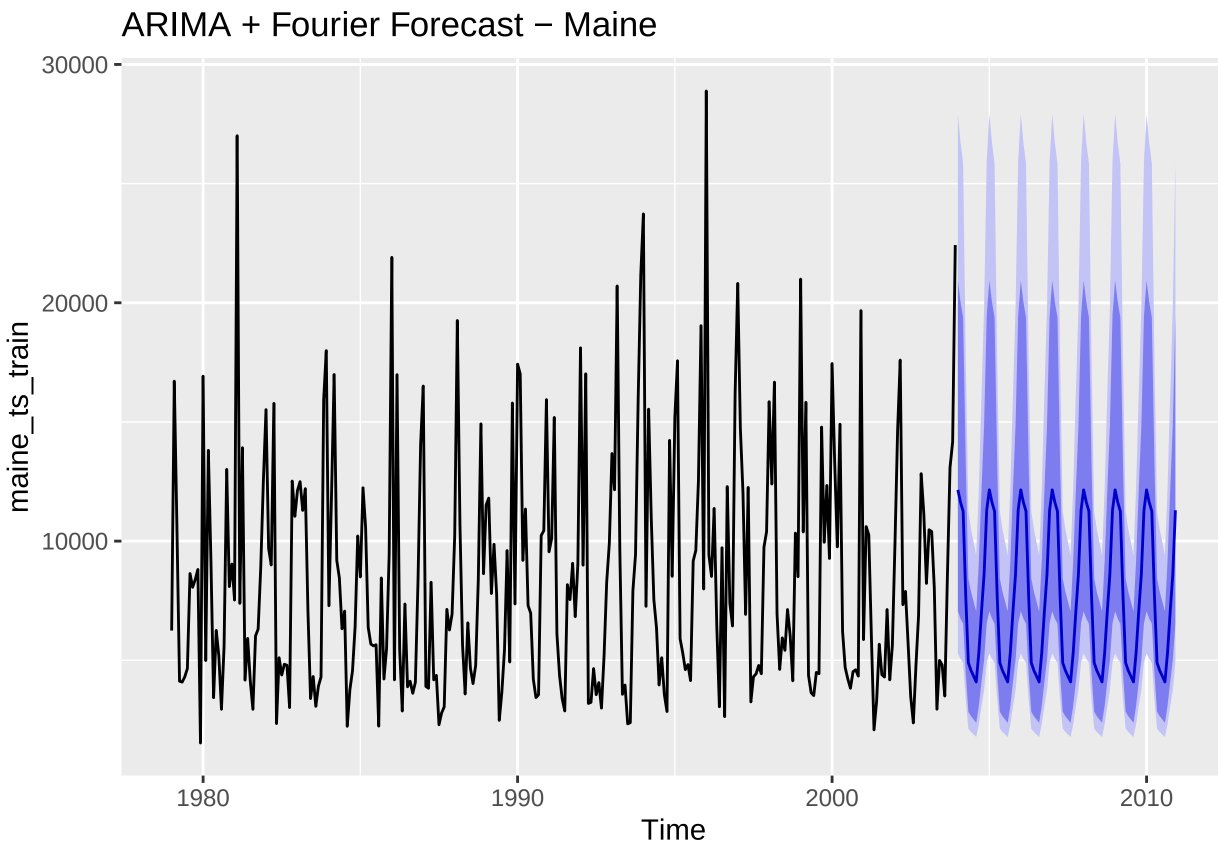
```
maine_sarima_fore <- forecast(object = maine_sarima_fit, h = fore_h)
autoplot(maine_sarima_fore) + ggtitle("SARIMA - Maine")
```



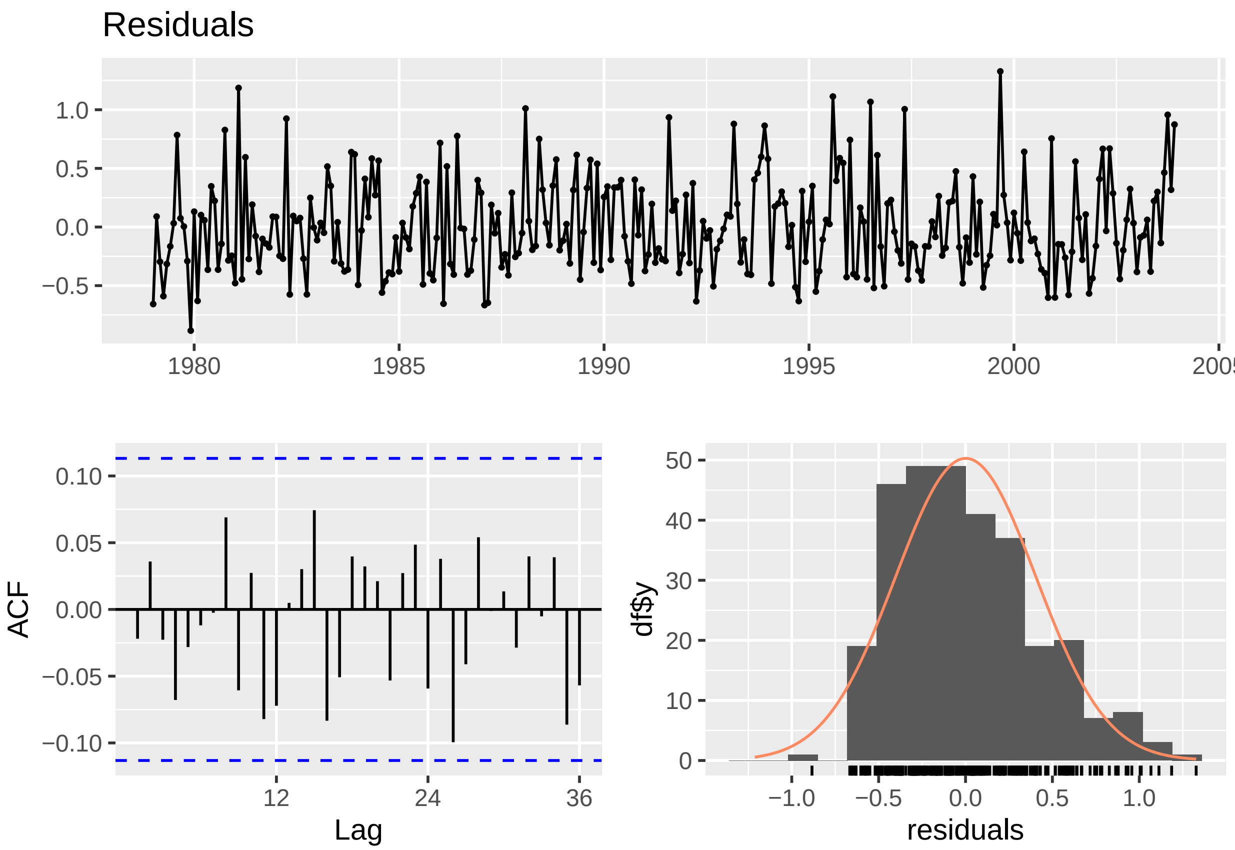

```r
#model 3: STL decomposition + ETS
maine_stlf_ets <- stlf(maine_ts_train, h = fore_h, method = "ets")
autoplot(maine_stlf_ets) + ggtitle("STL + ETS Forecast - Maine")
```




```r
#model 4: ARIMA + Fourier terms
k <- 4  #number of Fourier terms kept at 4, as 5 onwards, the RMSE started increasing

maine_arima_fit <- auto.arima(maine_ts_train,
                              seasonal=FALSE,
                              lambda=0,
                              xreg=fourier(maine_ts_train,
                                           K = k))

maine_arima_fore <- forecast(maine_arima_fit,
                            xreg=fourier(maine_ts_train,
                                         K = k,
                                         h = fore_h),
                            h = fore_h)
autoplot(maine_arima_fore) + ggtitle("ARIMA + Fourier Forecast - Maine")
```

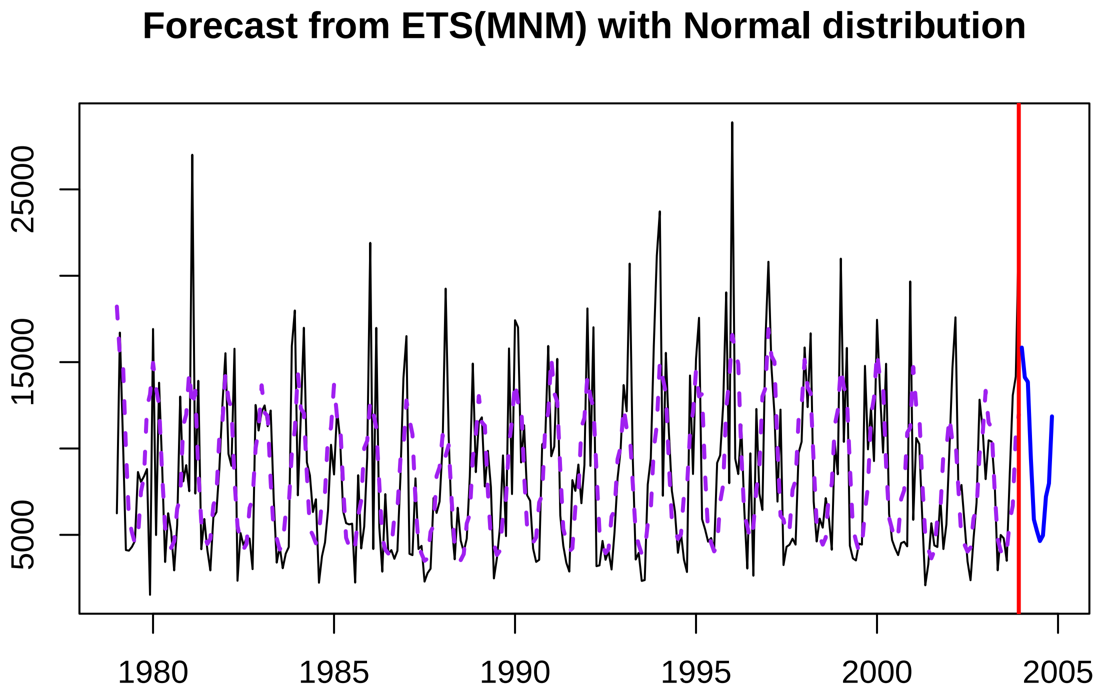



```r
n_for <- 11
#played with this, changed to 12, 23, 2 and check if the best model is still the same
maine_SSES <- es(maine_ts_train, model="ZZZ", h=n_for, holdout=FALSE)
checkresiduals(maine_SSES)
```




```
##
##  Ljung-Box test
##
## data:  Residuals
## Q* = 18.166, df = 24, p-value = 0.7949
##
## Model df: 0.   Total lags used: 24
```

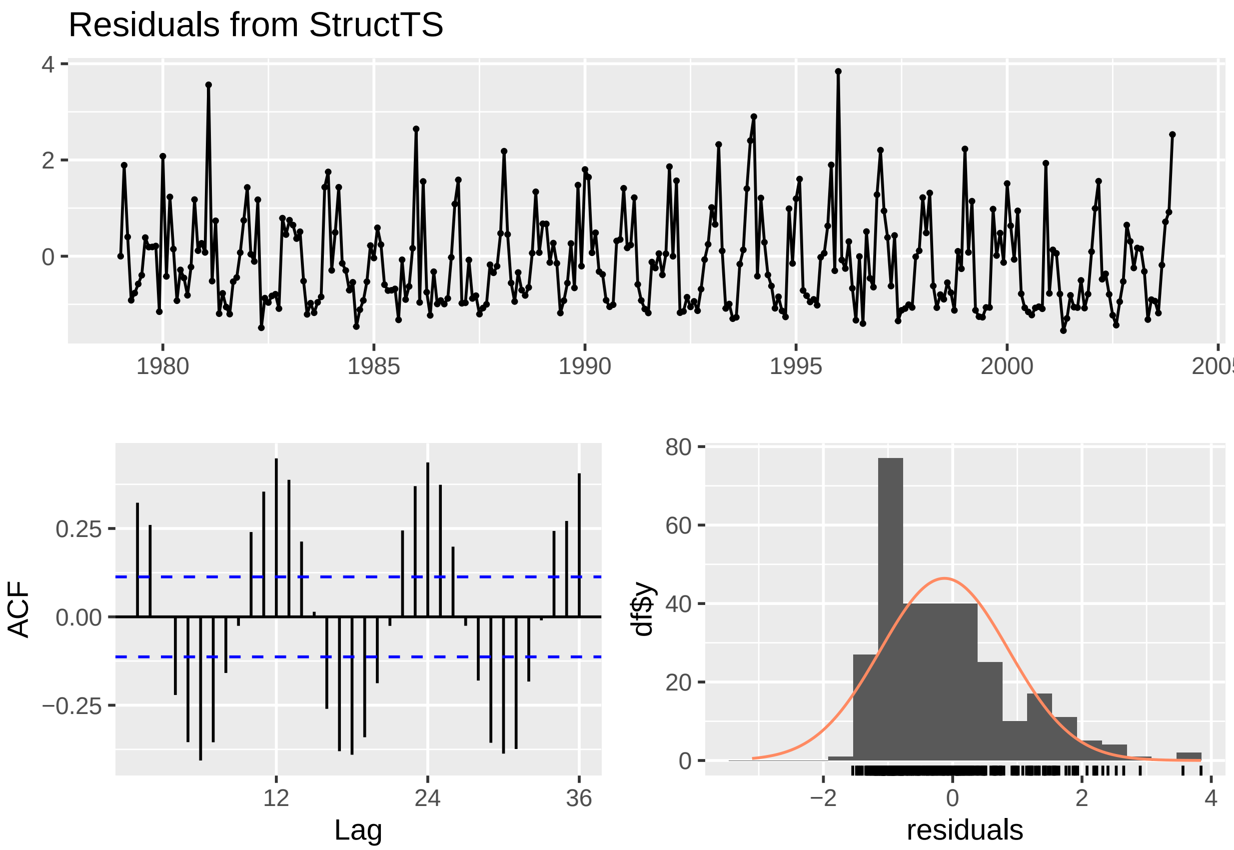
```
# forecast and plot
maine_SSES_fore <- forecast(maine_SSES,h=n_for)
plot(maine_SSES_fore)
```



Model intial interpretation: - Residuals kind of fluctuate around 0, but the variance is not very constant, but variance have slightly higher value fluctuations > 0, so some inconsistency with potential minor outliers. - ACF plot: no significant spikes outside of the blue lines, which is good - Histogram of residuals display rough normal distribution around 0, which is good. - Ljung-Box test: p-value (0.79) much greater than 0.05, so residuals are white noise, good.

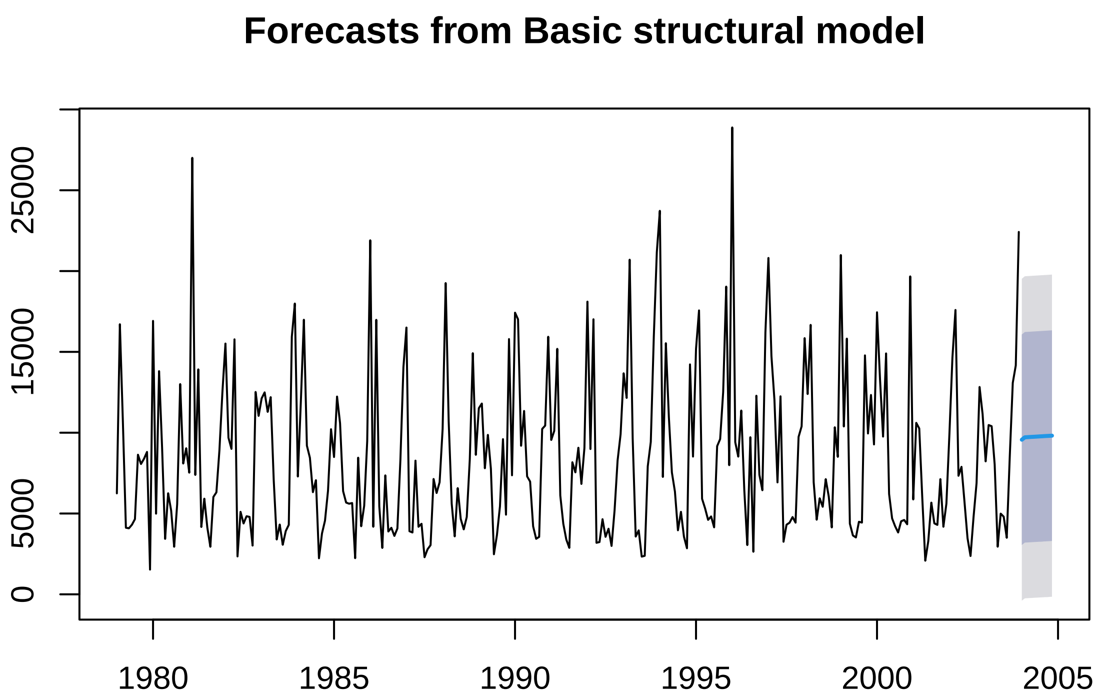
ACF: since the present value has very low correlation with the previous periods in the short term. Since we are looking at tidal power, longer-term trends may be more important, so es model did not do well.

```
SS_seas <- StructTS(maine_ts_train,
                    type="BSM",fixed=c(0.1,0.01,0.3,NA))

#this function has convergence issues
checkresiduals(SS_seas)
```




```
##
##  Ljung-Box test
##
## data:  Residuals from StructTS
## Q* = 678.76, df = 24, p-value < 2.2e-16
##
## Model df: 0.   Total lags used: 24
```

```
#Generating forecasts
maine_SS_for <- forecast(SS_seas,h=n_for)
plot(maine_SS_for)
```

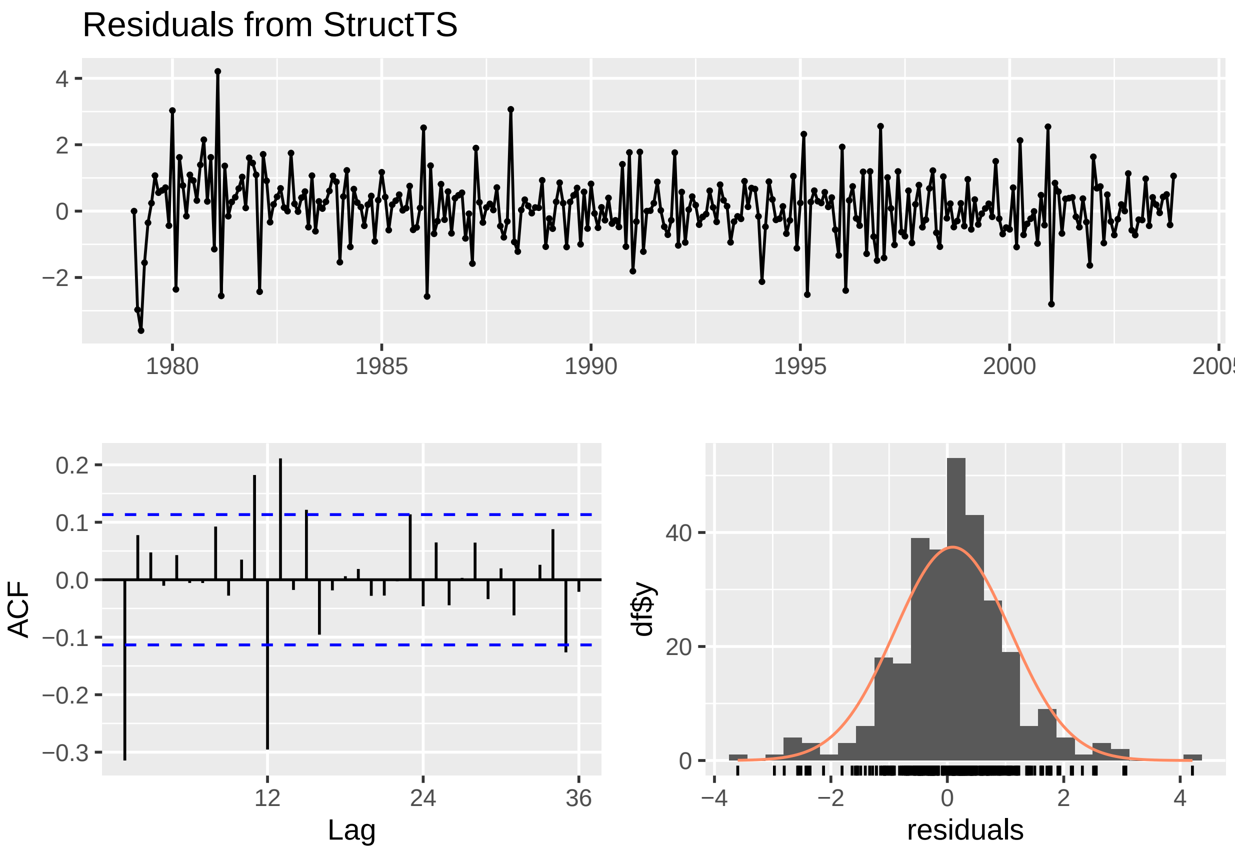


This model without changes from Module 10 was bad, strong autocorrelation and convergence issue, let's change:

```
maine_ts_train_diff <- diff(maine_ts_train, differences = 1)
SS_seas <- StructTS(maine_ts_train_diff, type = "BSM", , fixed = c(NA, NA, NA, NA))  # Let all paramete
```

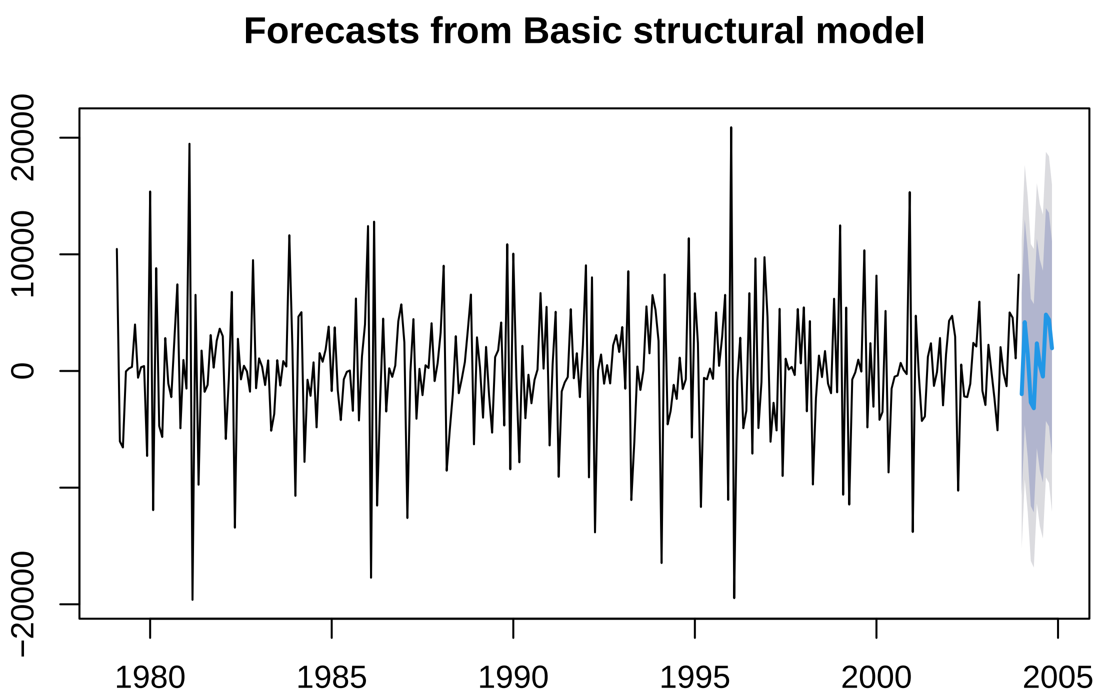
```
## Warning in StructTS(maine_ts_train_diff, type = "BSM", , fixed = c(NA, NA, :
## possible convergence problem: 'optim' gave code = 52 and message 'ERROR:
## ABNORMAL_TERMINATION_IN_LNSRCH'
```

```
checkresiduals(SS_seas)
```

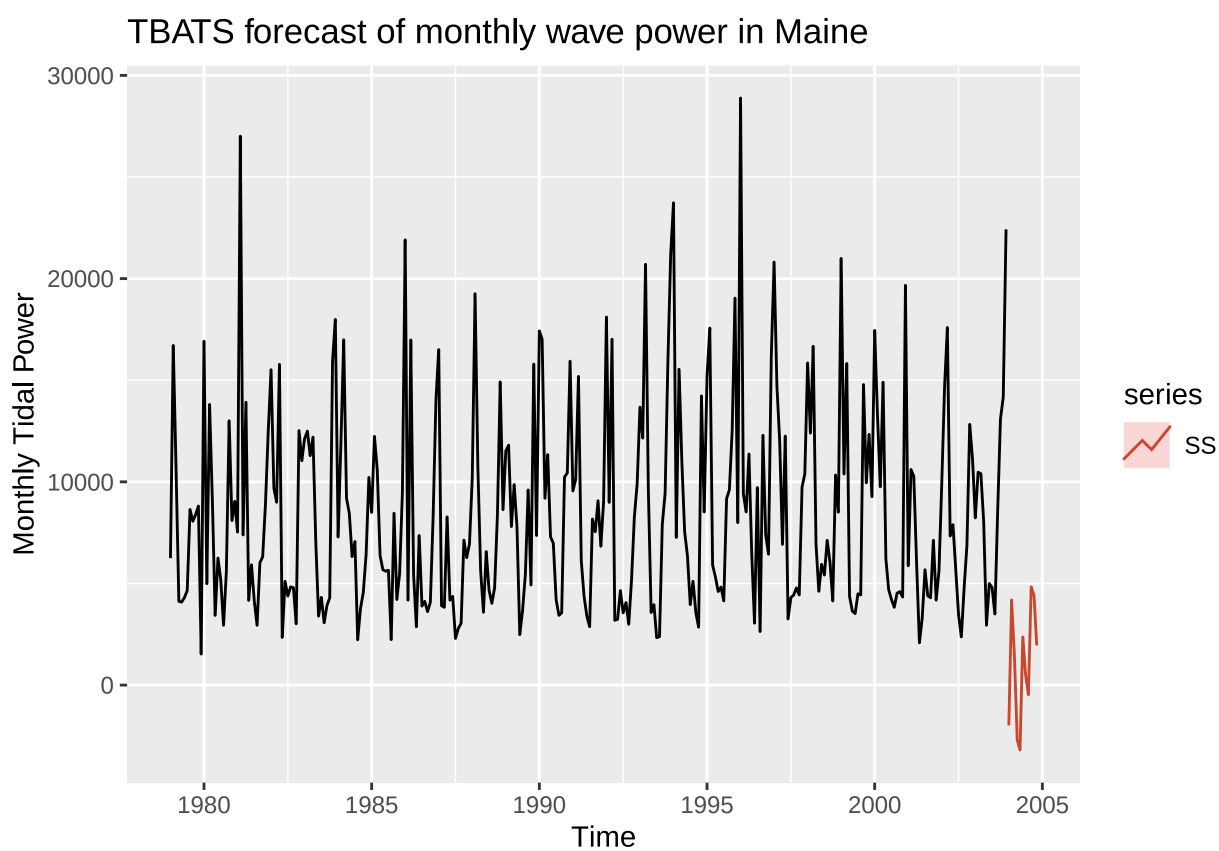



```
##
##  Ljung-Box test
##
## data:  Residuals from StructTS
## Q* = 101.35, df = 24, p-value = 1.764e-11
##
## Model df: 0.   Total lags used: 24
```

```
#Generating forecasts
maine_SS_for <- forecast(SS_seas,h=n_for)
plot(maine_SS_for)
```




```r
#Plot model + observed data
autoplot(maine_ts_train) +
  autolayer(maine_SS_for, series="SS",PI=FALSE)+
  ylab("Monthly Tidal Power") +
  ggtitle("TBATS forecast of monthly wave power in Maine")
```



Revised model: Residual plot improvements (no trends/outliers) suggest:

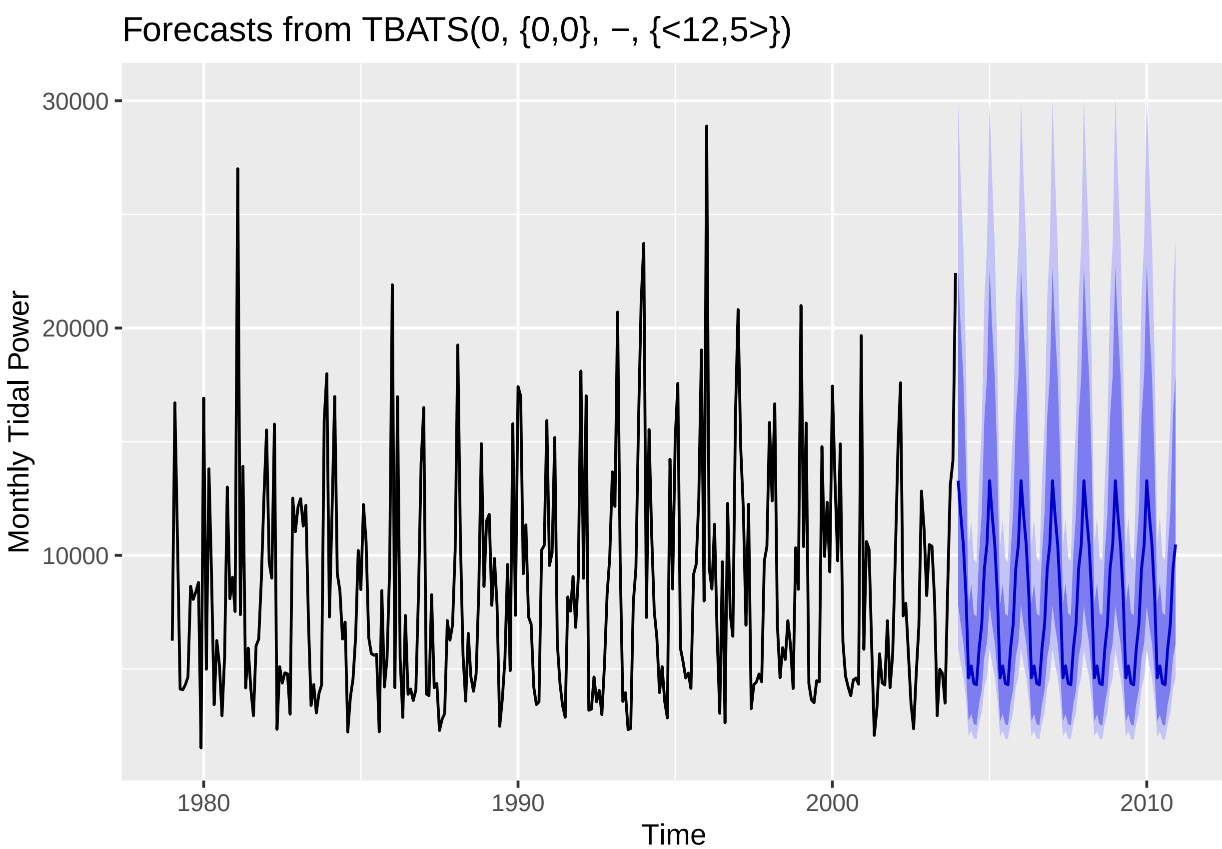
- The model handles mean and variance reasonably well.
- The seasonal/trend components are likely adequate.

ACF spikes + low p-value imply:

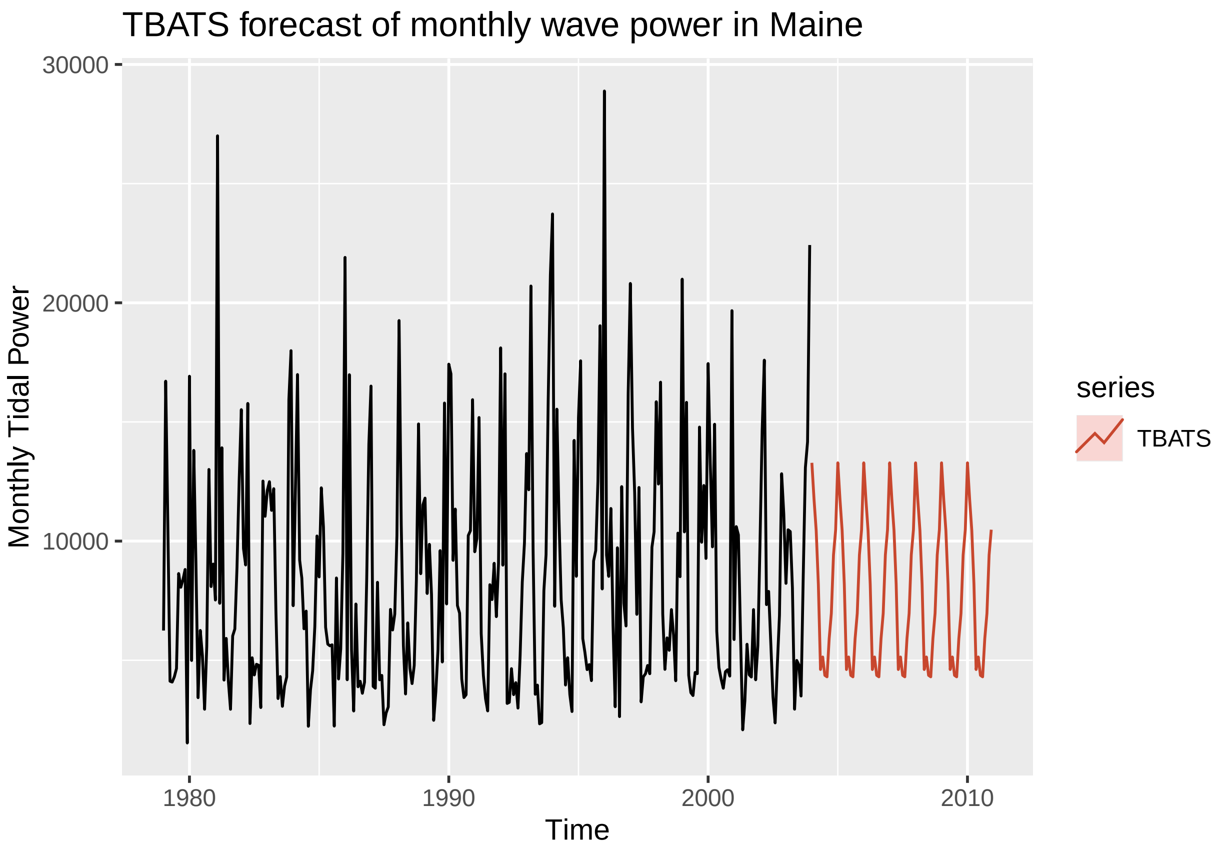
- Short-term dependencies remain unmodeled (e.g., AR/MA effects).
- Seasonal harmonics (higher-frequency cycles) may be missed.

Next steps could be: - combine the StructTS model with arima layer

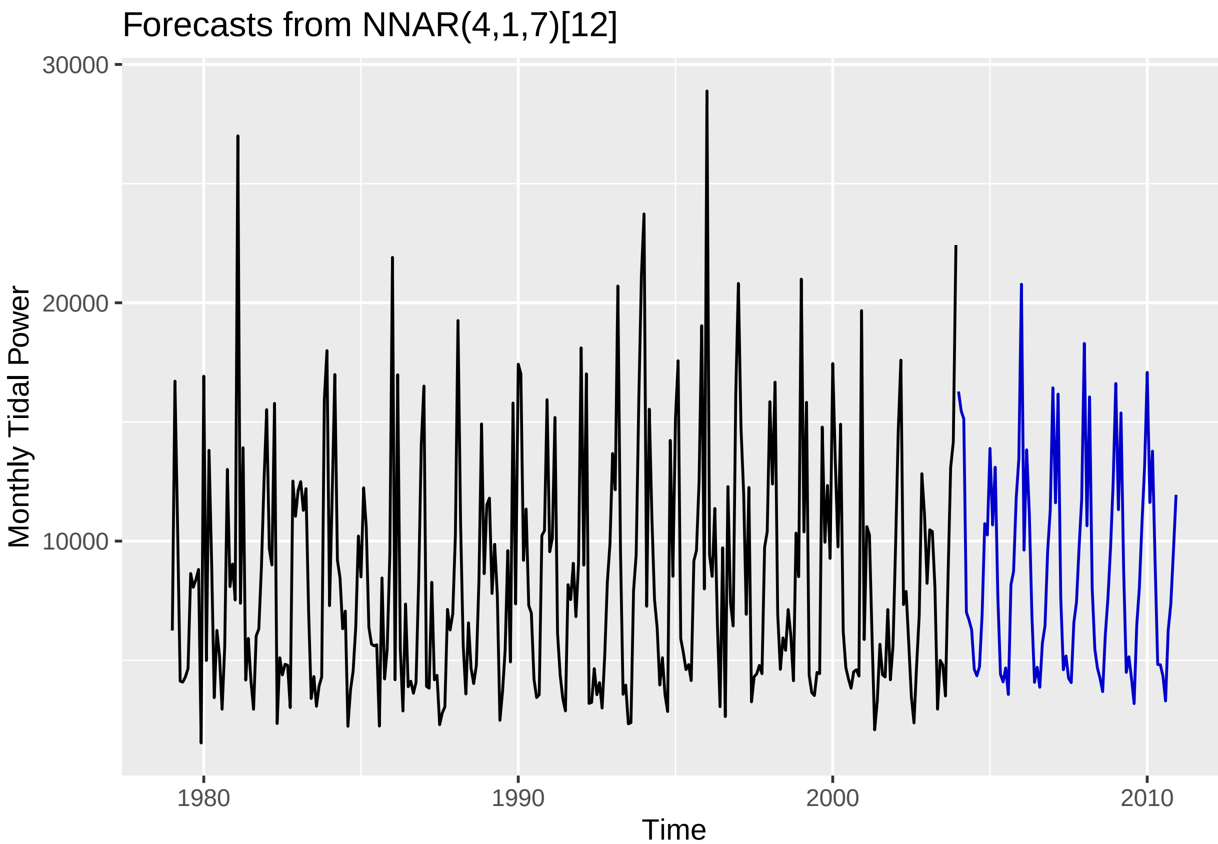
```r
#TBATS Model
maine_train_tbats_fit <- tbats(maine_ts_train)

tbats_forecast <- forecast(maine_train_tbats_fit, h=fore_h)

#Plot foresting results
autoplot(tbats_forecast) +
  ylab("Monthly Tidal Power")
```



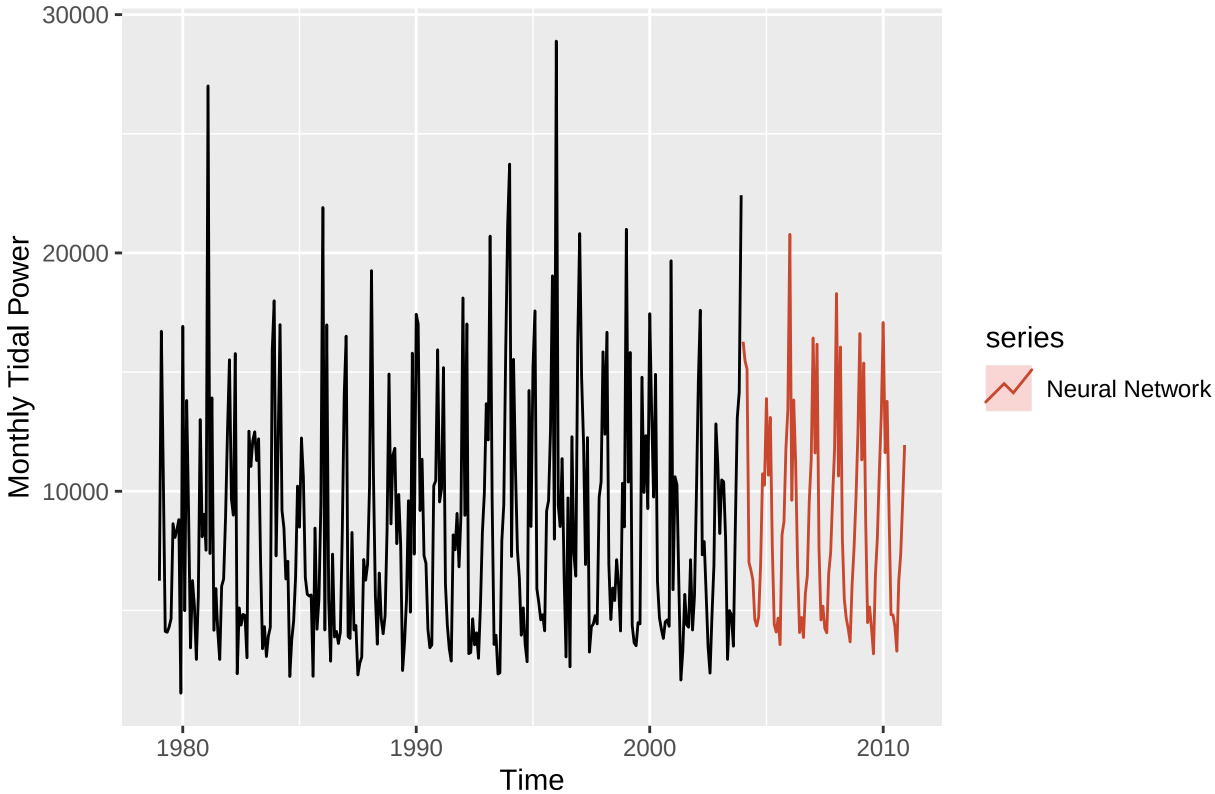

```r
#Plot model + observed data
autoplot(maine_ts_train) +
  autolayer(tbats_forecast, series="TBATS",PI=FALSE)+
  ylab("Monthly Tidal Power") +
  ggtitle("TBATS forecast of monthly wave power in Maine")
```




```
#NNETAR Model

maine_train_NN_fit <- nnetar(maine_ts_train,
                 p=4,
                 P=1,
                 xreg=fourier(maine_ts_train, K = 4))

NN_forecast <- forecast(maine_train_NN_fit,
                   h=fore_h,
                   xreg=fourier(maine_ts_train,
                                        K = 4, h=fore_h))


#Plot foresting results
autoplot(NN_forecast) +
  ylab("Monthly Tidal Power")
```




```r
#Plot model + observed data
autoplot(maine_ts_train) +
  autolayer(NN_forecast, series="Neural Network",PI=FALSE)+
  ylab("Monthly Tidal Power")
```



Observations: Both TBATS and NN under-forecast monthly tidal power. TBATS performs slightly better but both models do not seem to capture the seasonality of the monthly tidal power. For NN, some tweaking can be done to see if changing the lag can better capture the seasonality.

```
#summarizing the scores/results from the above models
SNAIVE_scores <- accuracy(maine_snaive$mean,maine_ts_test)
SARIMA_scores <- accuracy(maine_sarima_fore$mean,maine_ts_test)
ETS_scores <- accuracy(maine_stlf_ets$mean,maine_ts_test)
ARIMAF_scores <- accuracy(maine_arima_fore$mean,maine_ts_test)
es_scores <- accuracy(maine_SS_for$mean,maine_ts_test)
StructTS_scores <- accuracy(maine_SS_for$mean,maine_ts_test)
TBATS_scores <- accuracy(tbats_forecast$mean,maine_ts_test)
NN_scores <- accuracy(NN_forecast$mean,maine_ts_test)

scores <- as.data.frame(
  rbind(SNAIVE_scores, SARIMA_scores, ETS_scores, ARIMAF_scores, es_scores, StructTS_scores,
        TBATS_scores, NN_scores)
  )
row.names(scores) <- c("SNAIVE", "SARIMA", "STL+ETS", "ARIMA+Fourier",
                       "ES", "StructTS", "TBAT", "NN")

#choose model with lowest RMSE
best_model_index <- which.min(scores[,"RMSE"])
cat("The best model by RMSE is:", row.names(scores[best_model_index,]))
```

```
## The best model by RMSE is: ARIMA+Fourier
```

```
kbl(scores,
      caption = "Forecast Accuracy for Monthly Wave Power - Maine",
      digits = array(5,ncol(scores))) %>%
```

```r
  kable_styling(full_width = FALSE, position = "center", latex_options = "hold_position") %>%
  #highlight model with lowest RMSE
  kable_styling(latex_options="striped", stripe_index = which.min(scores[,"RMSE"]))
```

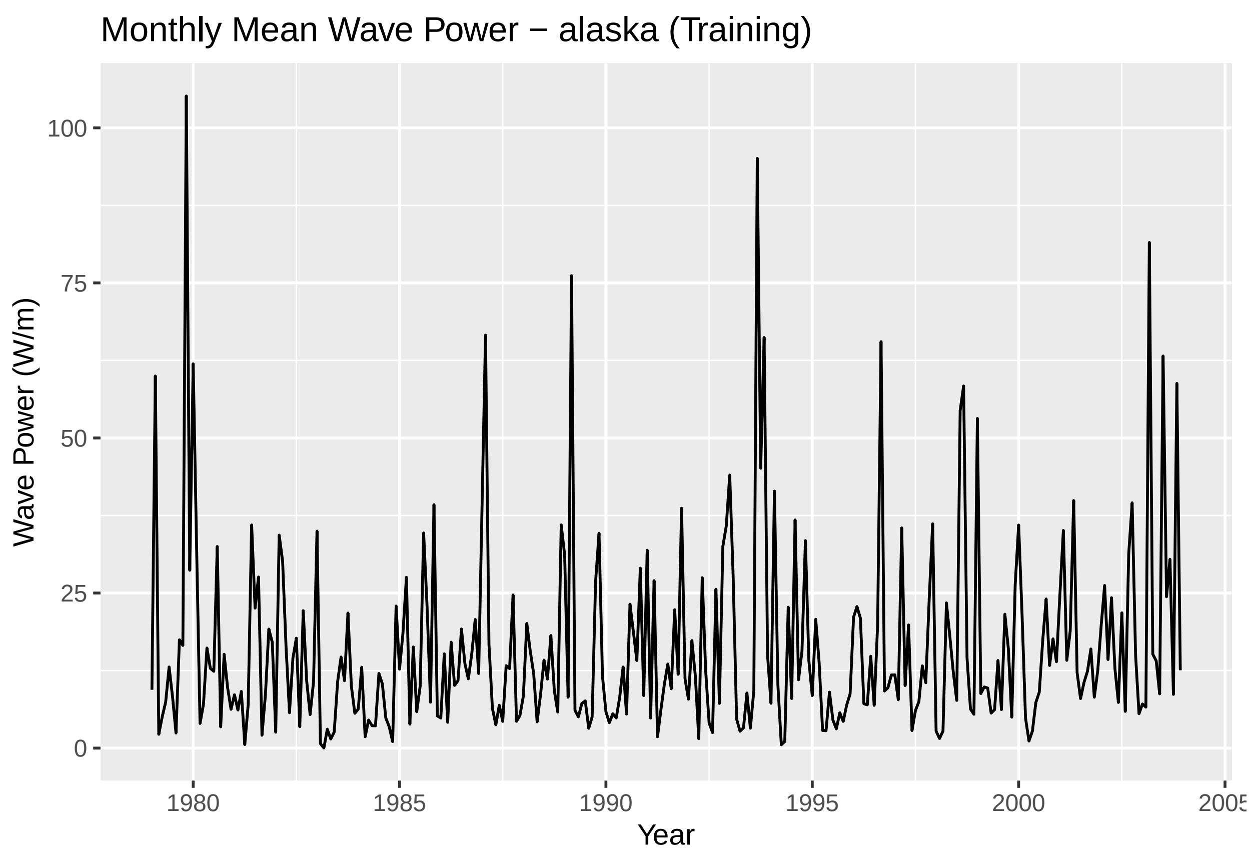
Table 1: Forecast Accuracy for Monthly Wave Power - Maine

	ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's U
SNAIVE	-822.5207	4290.692	3202.526	-19.62825	42.60721	0.06549	0.89498
SARIMA	-217.5947	4026.117	3221.739	-29.70686	50.01054	0.32049	0.83462
STL+ETS	-192.9789	3730.735	2709.232	-17.21843	36.39588	0.18831	0.70042
ARIMA+Fourier	736.3832	3690.025	2601.569	-5.30294	31.99013	0.20083	0.72561
ES	4686.0140	5864.526	4762.718	78.82107	80.54519	-0.00122	1.84991
StructTS	4686.0140	5864.526	4762.718	78.82107	80.54519	-0.00122	1.84991
TBAT	569.0860	3728.780	2690.142	-8.37019	33.99924	0.17281	0.70592
NN	-388.5744	4254.848	3093.926	-18.16257	40.01905	0.11345	0.79176

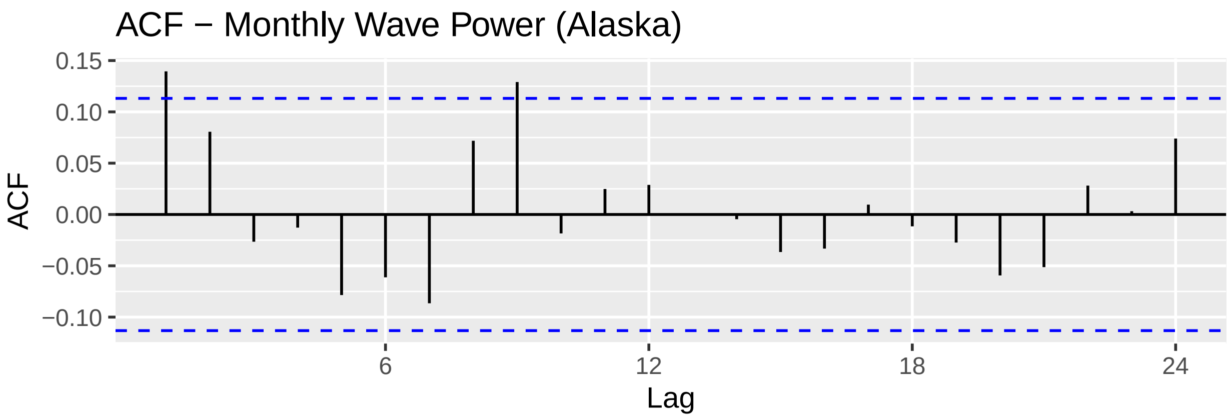
#Alaska

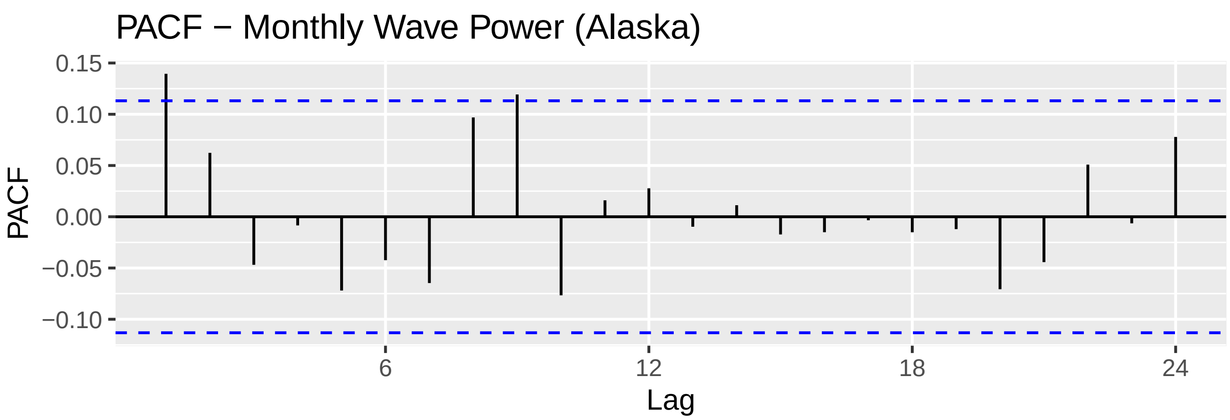
###Start by creating monthly time series objects for Alaska and plotting ACF and PACF plots

```r
#converting to time series object
alaska_ts_train <- ts(alaska_monthly_train_df$monthly_mean_wave_power,
                      start = c(min(alaska_monthly_train_df$Year), min(alaska_monthly_train_df$Month)),
                      frequency = 12)

alaska_ts_test <- ts(alaska_monthly_test_df$monthly_mean_wave_power,
                      start = c(min(alaska_monthly_test_df$Year), min(alaska_monthly_test_df$Month)),
                      frequency = 12)

alaska_train_plot <- autoplot(alaska_ts_train) +
  ggtitle("Monthly Mean Wave Power - alaska (Training)") +
  xlab("Year") + ylab("Wave Power (W/m)")
plot(alaska_train_plot)
```

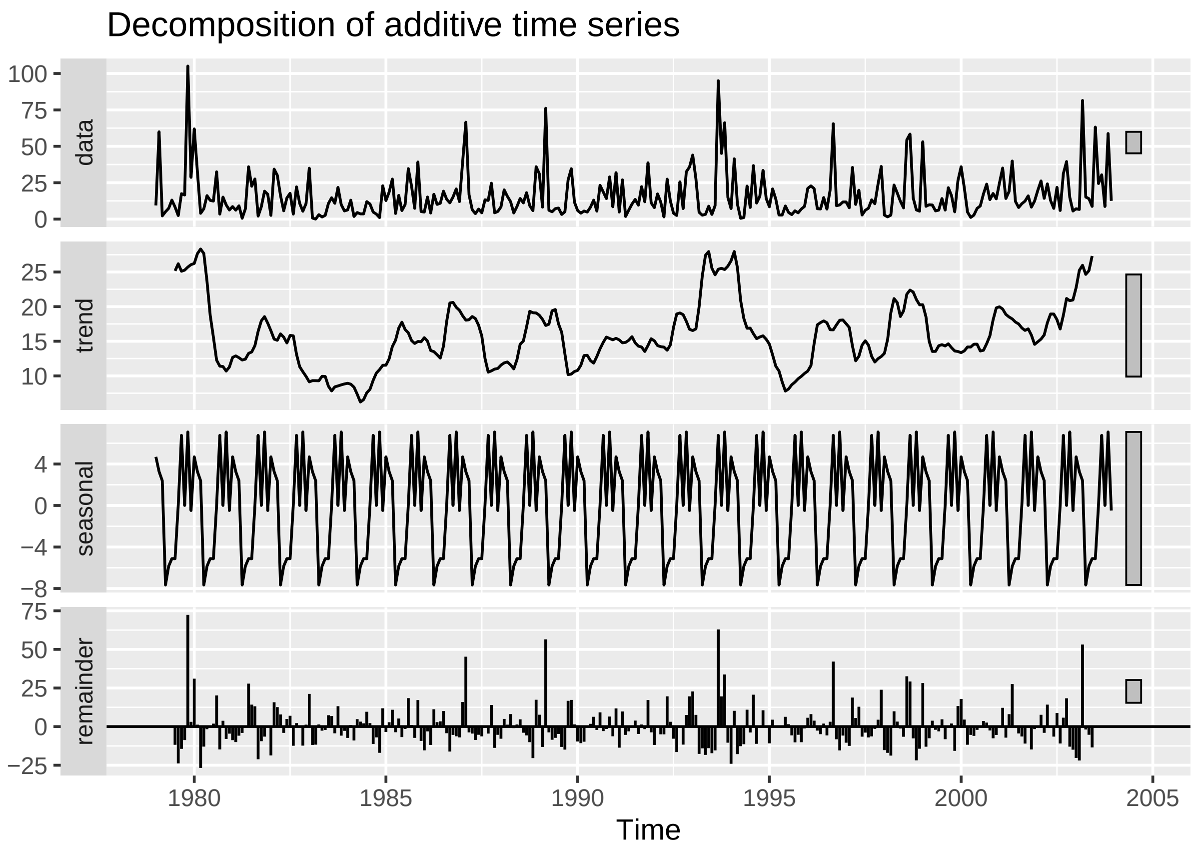



```r
#plotting acf and pacf
alaska_train_acf <- ggAcf(alaska_ts_train) + ggtitle("ACF - Monthly Wave Power (Alaska)")
alaska_train_pacf <- ggPacf(alaska_ts_train) + ggtitle("PACF - Monthly Wave Power (Alaska)")
plot_grid(alaska_train_acf, alaska_train_pacf, ncol = 1)
```






```r
#decomposing the time series
alaska_train_decomp <- decompose(alaska_ts_train)
autoplot(alaska_train_decomp)
```



Observations: The significant spike at lag 1 in both ACF and PACF strongly suggests an AR(1) component. Also, the significant spikes at lag 12 in both ACF and PACF indicate a strong seasonal autoregressive component with a period of 12 months (SAR(1) with a seasonal lag of 12).

```
adf_alaska <- adf.test(alaska_ts_train)
```

```
## Warning in adf.test(alaska_ts_train): p-value smaller than printed p-value
```

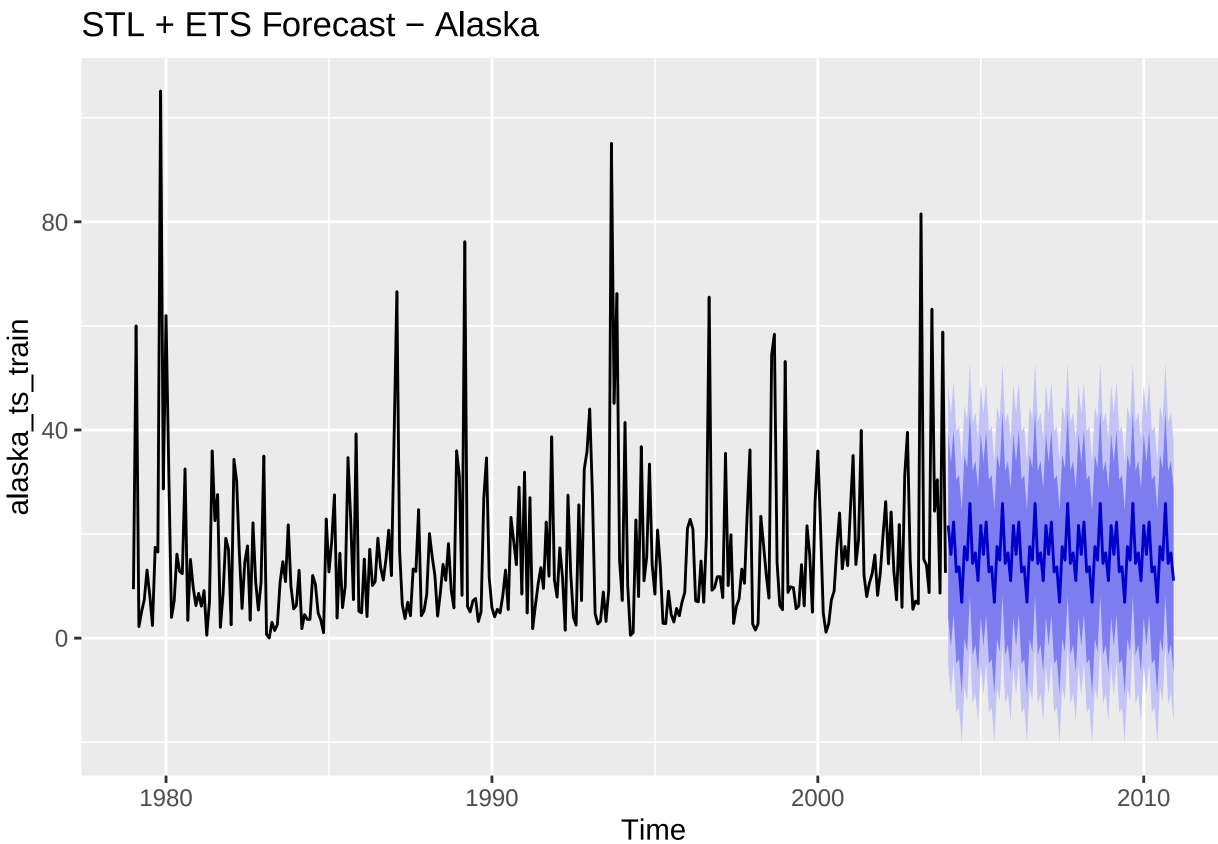
```
# Print the results
print(adf_alaska)
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  alaska_ts_train
## Dickey-Fuller = -7.3596, Lag order = 6, p-value = 0.01
## alternative hypothesis: stationary
```

Observations: ADF test returns a p-value of 0.01, which is smaller than our chosen significance level of 0.05. We reject the null hypothesis that the Alaska mean monthly wave power series has a unit root and thus, the series is likely stationary and does not need differencing.

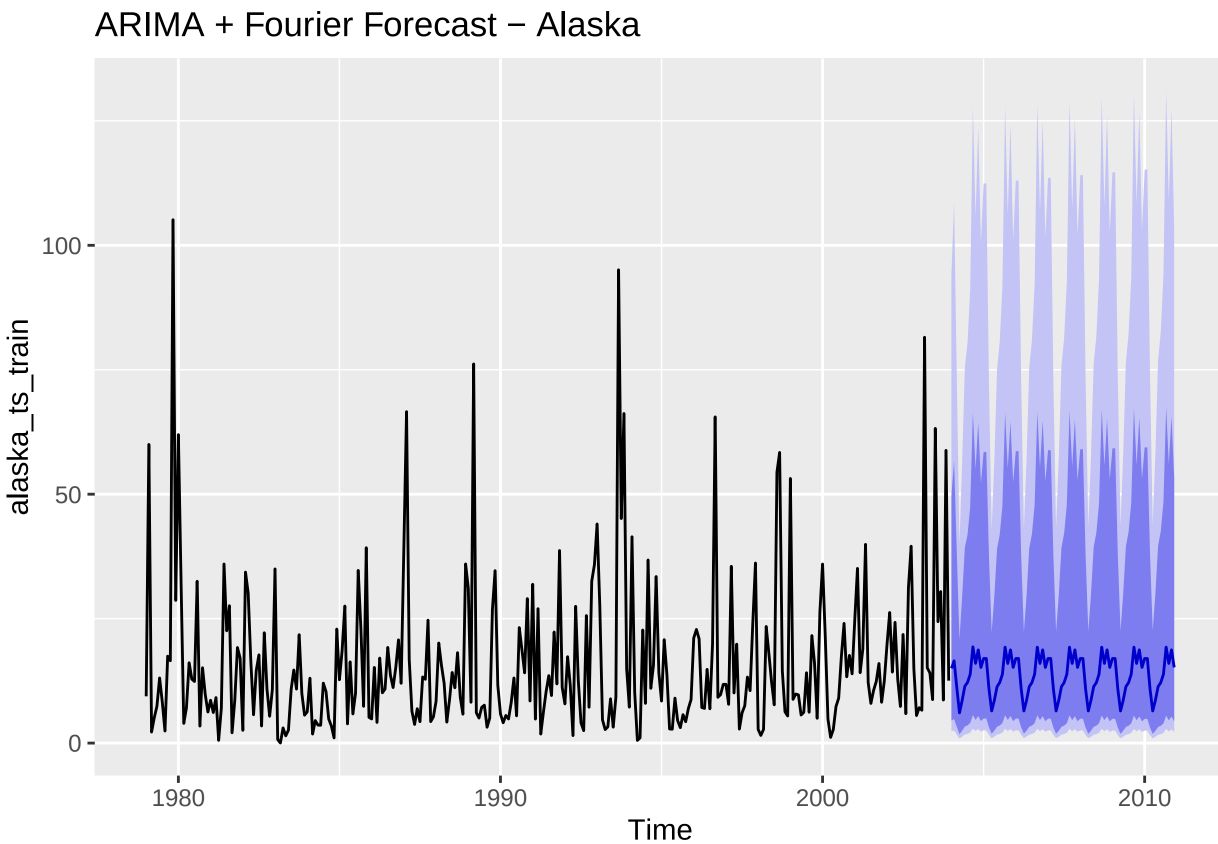
###Proceed with using our 3 chosen models on Alaska ## Model 1: STL decomposition + ETS

```
fore_h <- length(alaska_ts_test)

#model 1: STL decomposition + ETS
alaska_stlf_ets <- stlf(alaska_ts_train, h = fore_h, method = "ets")
autoplot(alaska_stlf_ets) + ggtitle("STL + ETS Forecast - Alaska")
```



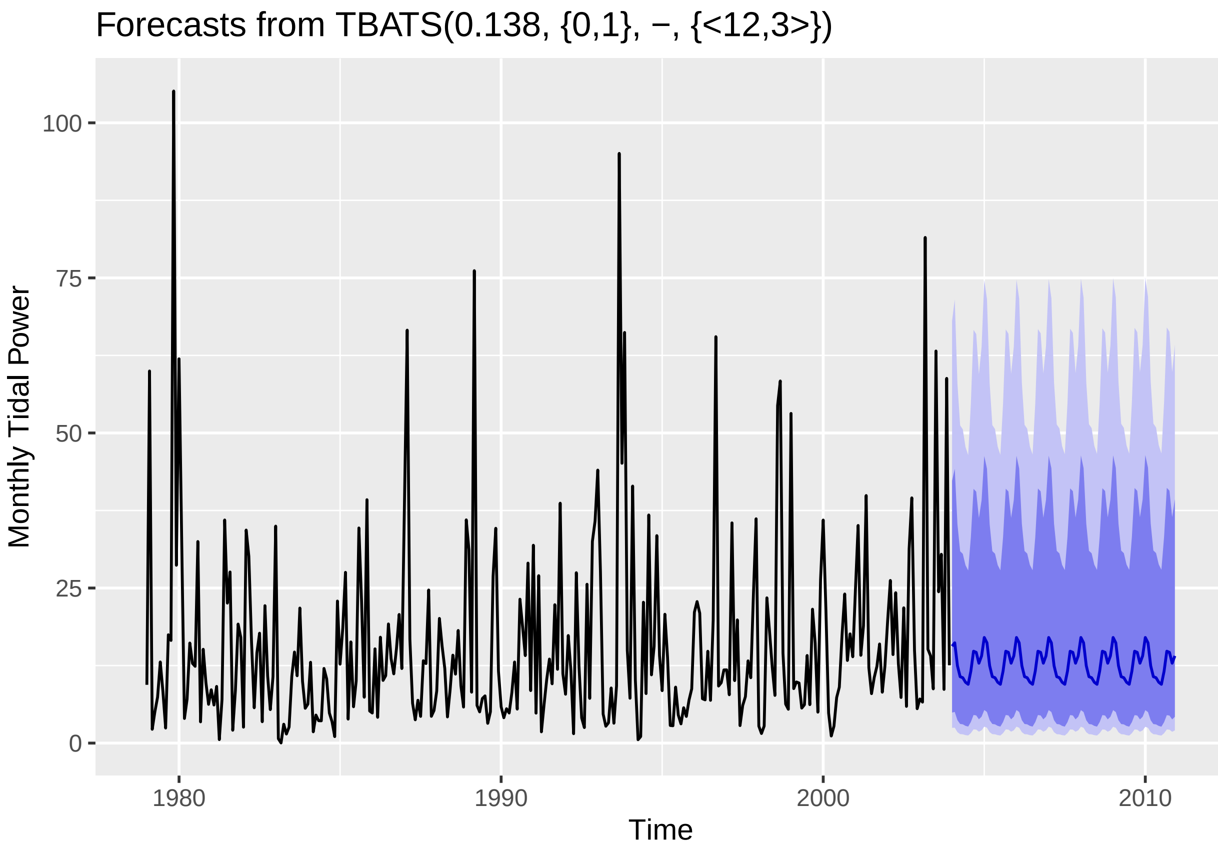
Model 2: ARIMA + Fourier terms

```
#Model 2: Arima + Fourier
k <- 6 #Value that gives lowest RMSE - need to justify
alaska_arima_fit <- auto.arima(alaska_ts_train,
                               seasonal=FALSE,
                               lambda=0,
                               xreg=fourier(alaska_ts_train,
                                            K = k))

alaska_arima_fore <- forecast(alaska_arima_fit,
                              xreg=fourier(alaska_ts_train,
                                           K = k,
                                           h = fore_h),
                              h = fore_h)
autoplot(alaska_arima_fore) + ggtitle("ARIMA + Fourier Forecast - Alaska")
```

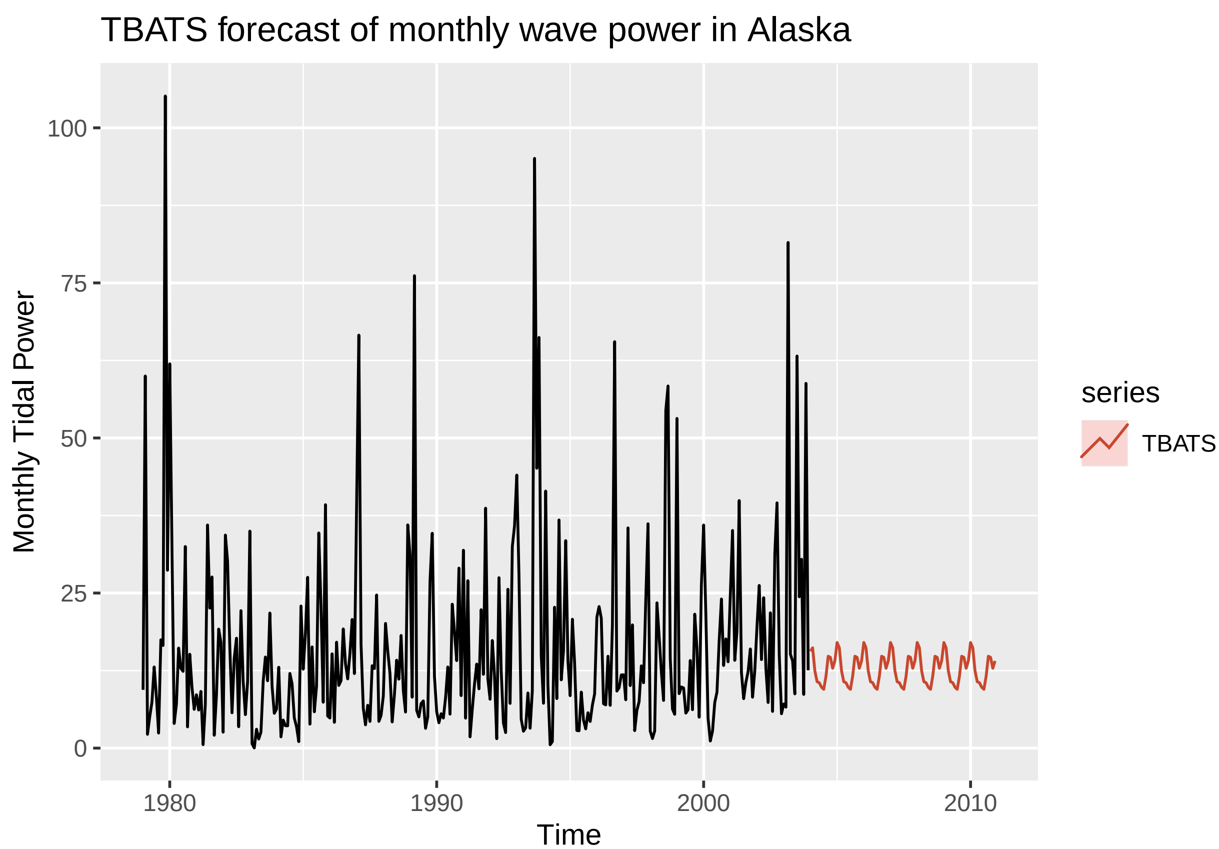


Model 3: TBATS

```
#Model 3: TBATS
alaska_train_tbats_fit <- tbats(alaska_ts_train)

tbats_forecast <- forecast(alaska_train_tbats_fit, h=fore_h)

#Plot foresting results
autoplot(tbats_forecast) +
  ylab("Monthly Tidal Power")
```




```r
#Plot model + observed data
autoplot(alaska_ts_train) +
  autolayer(tbats_forecast, series="TBATS",PI=FALSE)+
  ylab("Monthly Tidal Power") +
  ggtitle("TBATS forecast of monthly wave power in Alaska")
```




```
#summarizing the scores/results from the above models
ETS_scores <- accuracy(alaska_stlf_ets$mean,alaska_ts_test)
ARIMAF_scores <- accuracy(alaska_arima_fore$mean,alaska_ts_test)
TBATS_scores <- accuracy(tbats_forecast$mean,alaska_ts_test)

scores <- as.data.frame(
  rbind(ETS_scores, ARIMAF_scores, TBATS_scores)
  )
row.names(scores) <- c("STL+ETS", "ARIMA+Fourier", "TBAT")

#choose model with lowest RMSE
best_model_index <- which.min(scores[,"RMSE"])
cat("The best model by RMSE is:", row.names(scores[best_model_index,]))
```

```
## The best model by RMSE is: ARIMA+Fourier
```

```
kbl(scores,
      caption = "Forecast Accuracy for Monthly Wave Power - alaska",
      digits = array(5,ncol(scores))) %>%
  kable_styling(full_width = FALSE, position = "center", latex_options = "hold_position") %>%
  #highlight model with lowest RMSE
  kable_styling(latex_options="striped", stripe_index = which.min(scores[,"RMSE"]))
```

#Florida

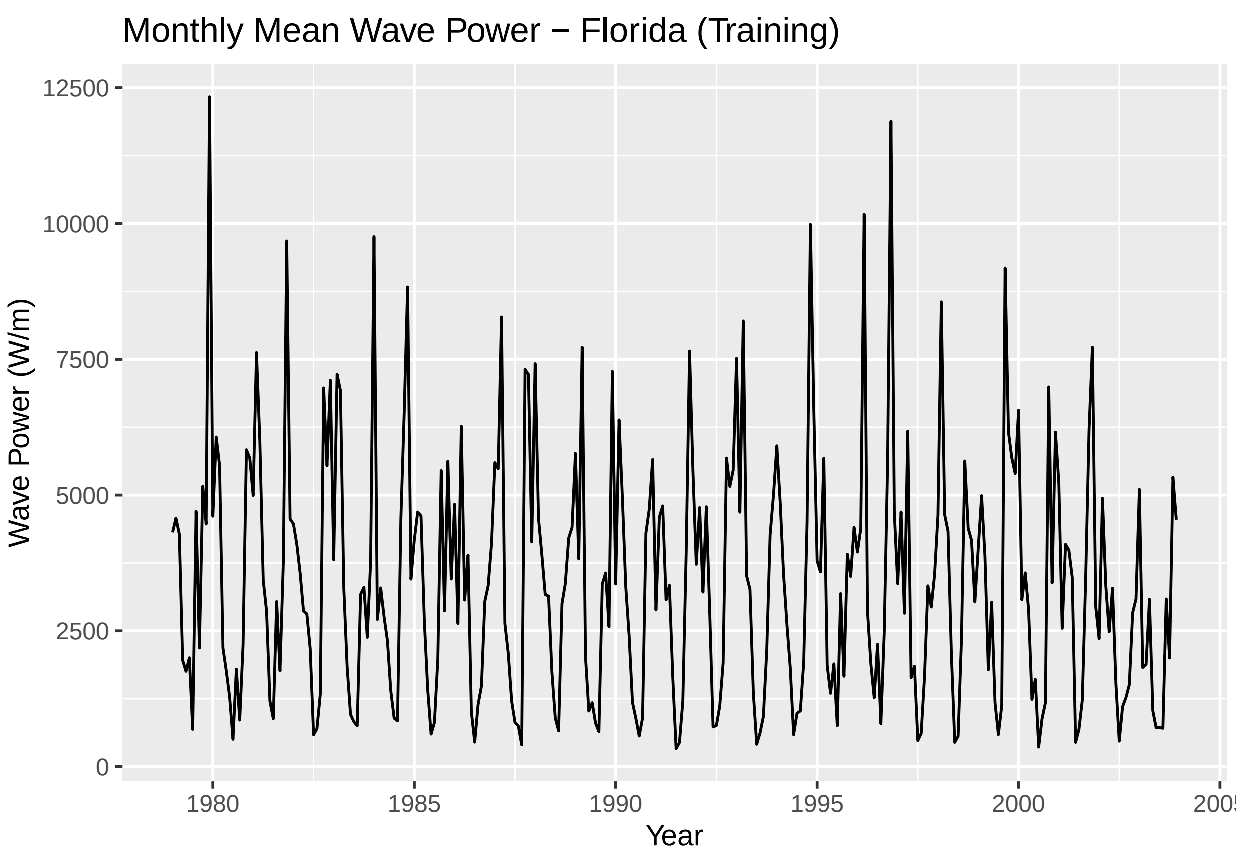
###Start by creating monthly time series objects for Florida and plotting ACF and PACF plots

```
#converting to time series object
florida_ts_train <- ts(florida_monthly_train_df$monthly_mean_wave_power,
```

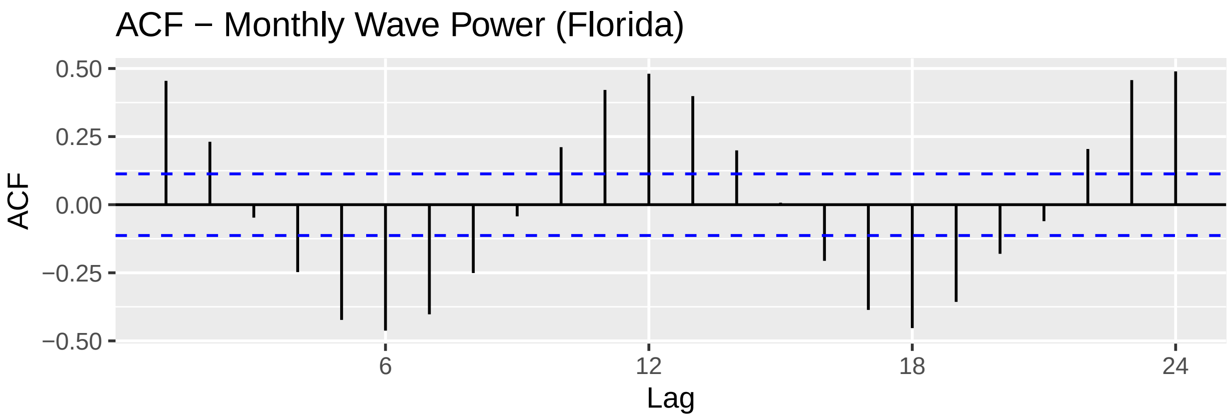
Table 2: Forecast Accuracy for Monthly Wave Power - alaska

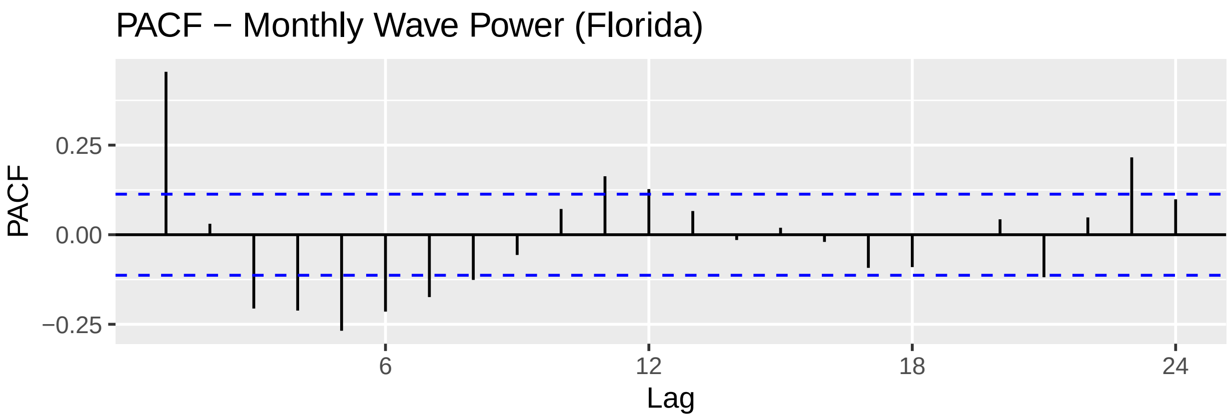
	ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's U
STL+ETS	0.41065	14.12462	10.06050	-118.37052	148.4886	0.24409	0.44005
ARIMA+Fourier	2.70368	13.68095	9.00760	-79.91860	115.1316	0.19990	0.39252
TBAT	3.71047	14.20461	8.81815	-71.65192	107.4045	0.18163	0.41341

```
                      start = c(min(florida_monthly_train_df$Year), min(florida_monthly_train_df$Month))
                      frequency = 12)

florida_ts_test <- ts(florida_monthly_test_df$monthly_mean_wave_power,
                      start = c(min(florida_monthly_test_df$Year), min(florida_monthly_test_df$Month)),
                      frequency = 12)

florida_train_plot <- autoplot(florida_ts_train) +
  ggtitle("Monthly Mean Wave Power - Florida (Training)") +
  xlab("Year") + ylab("Wave Power (W/m)")
plot(florida_train_plot)
```

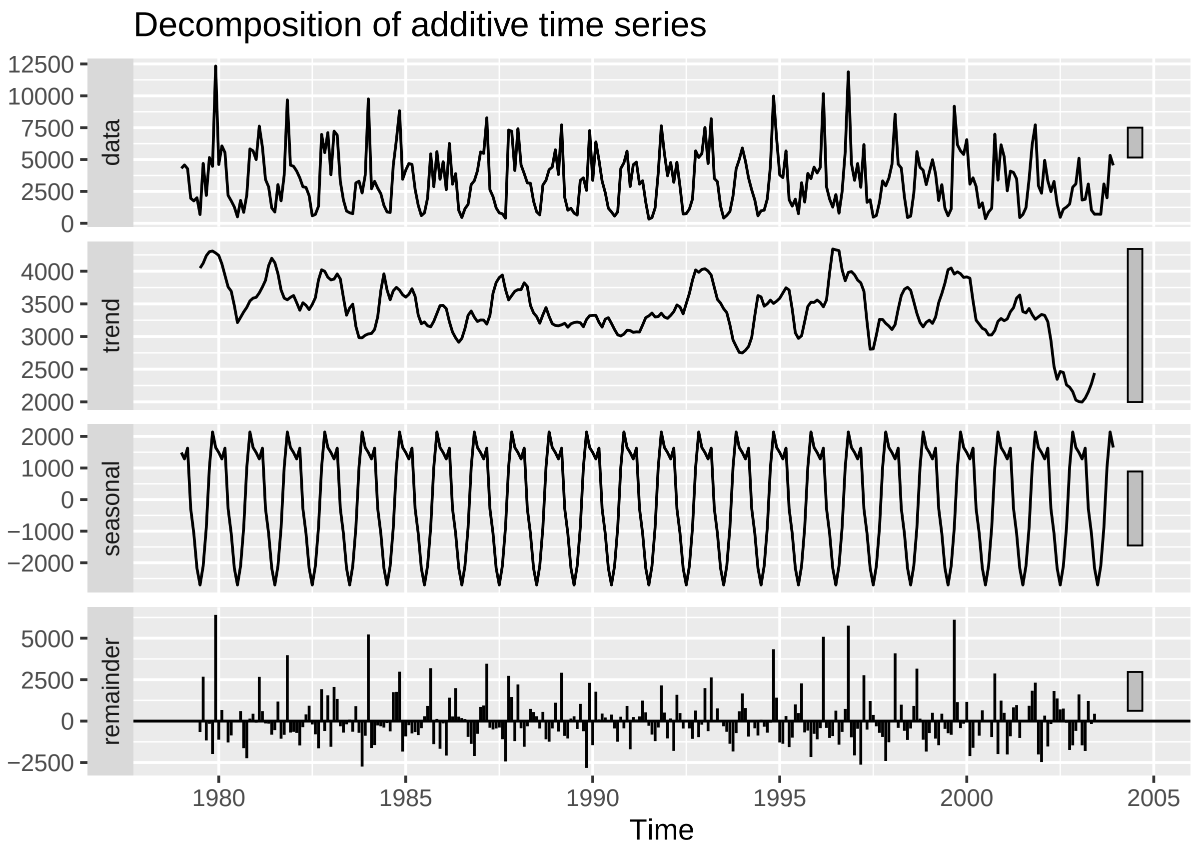



```
#plotting acf and pacf
florida_train_acf <- ggAcf(florida_ts_train) + ggtitle("ACF - Monthly Wave Power (Florida)")
florida_train_pacf <- ggPacf(florida_ts_train) + ggtitle("PACF - Monthly Wave Power (Florida)")
plot_grid(florida_train_acf, florida_train_pacf, ncol = 1)
```






```r
#decomposing the time series
florida_train_decomp <- decompose(florida_ts_train)
autoplot(florida_train_decomp)
```



Observations: The significant spike at lag 1 in both ACF and PACF strongly suggests an AR(1) component. Also, the repeating seasonal patterns at lag 12 for the ACF suggest strong yearly seasonality.

```
adf_florida <- adf.test(florida_ts_train)
```

```
## Warning in adf.test(florida_ts_train): p-value smaller than printed p-value
```

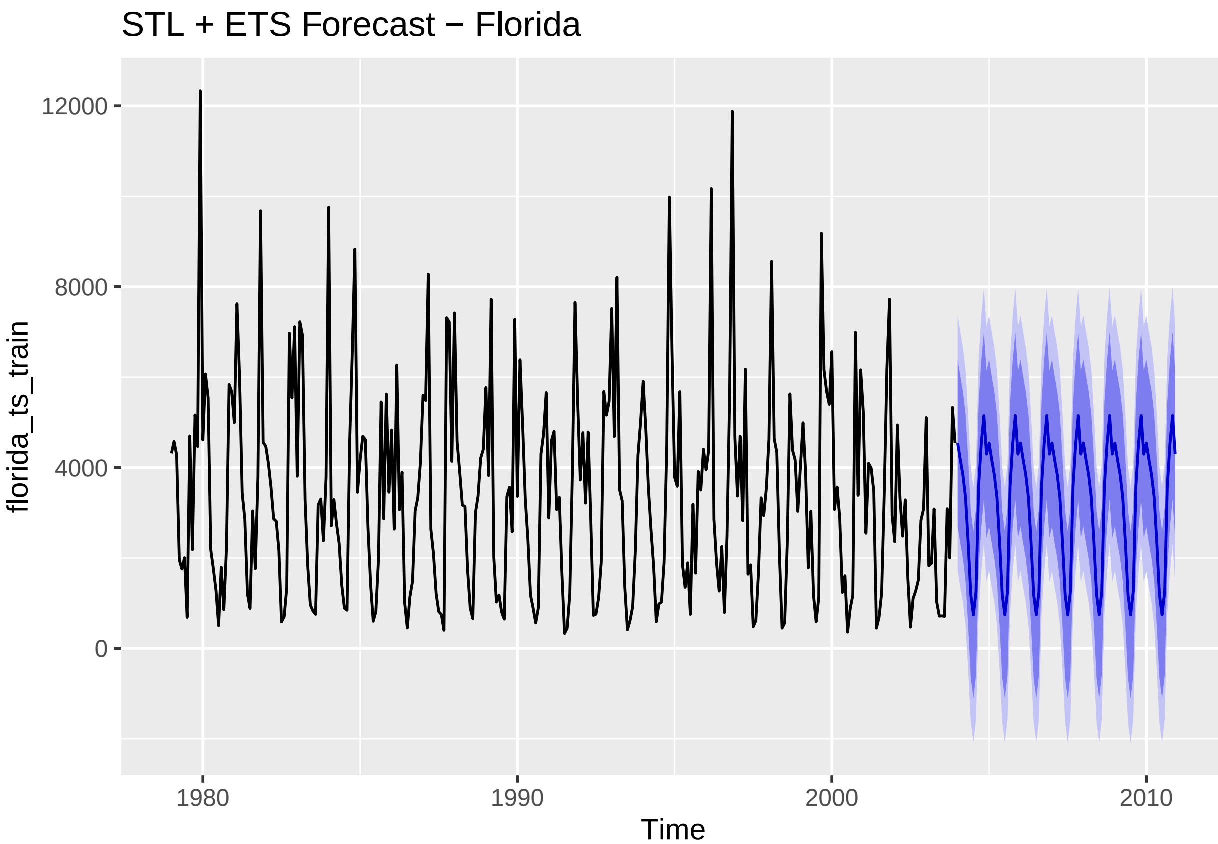
```
# Print the results
print(adf_florida)
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  florida_ts_train
## Dickey-Fuller = -12.062, Lag order = 6, p-value = 0.01
## alternative hypothesis: stationary
```

Observations: ADF test returns a p-value of 0.01, which is smaller than our chosen significance level of 0.05. We reject the null hypothesis that the Alaska mean monthly wave power series has a unit root and thus, the series is likely stationary and does not need differencing.

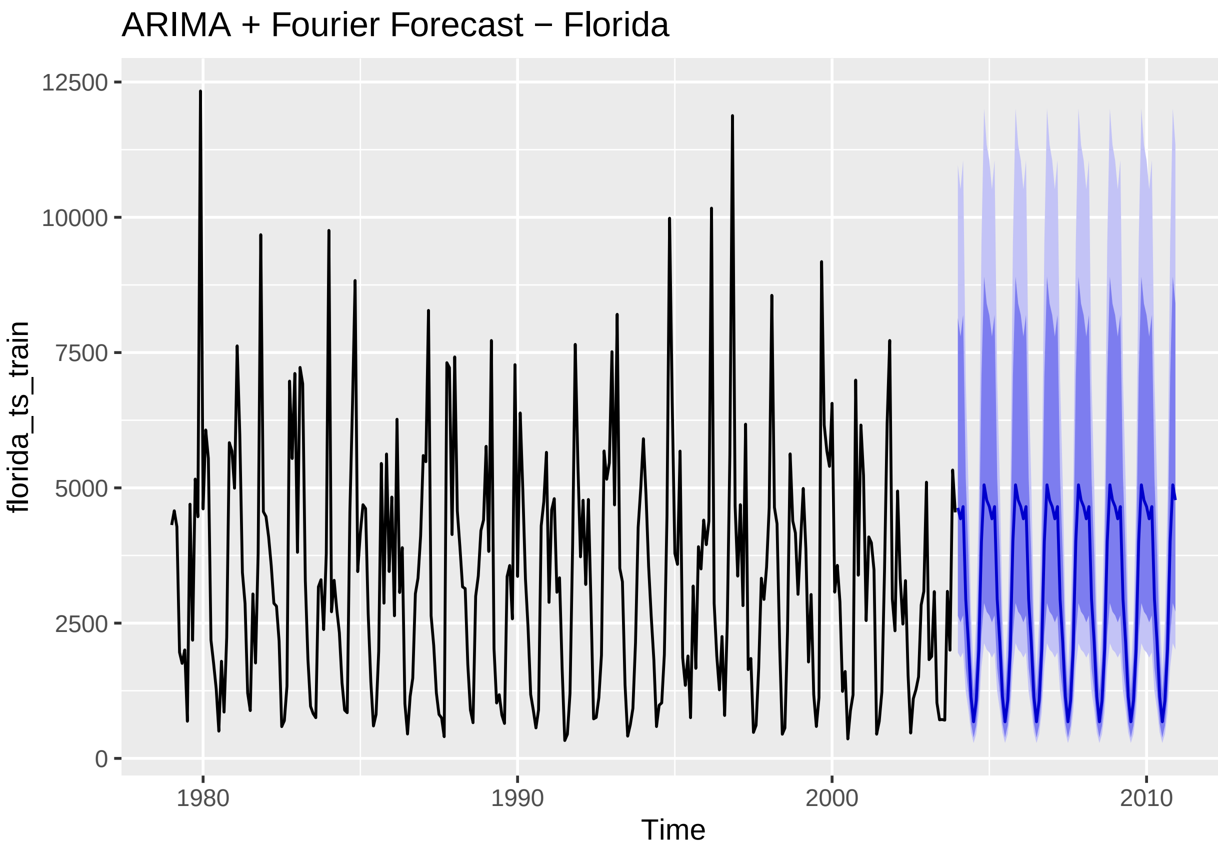
###Proceed with using our 3 chosen models on Florida

```
fore_h <- length(florida_ts_test)

#model 1: STL decomposition + ETS
florida_stlf_ets <- stlf(florida_ts_train, h = fore_h, method = "ets")
autoplot(florida_stlf_ets) + ggtitle("STL + ETS Forecast - Florida")
```



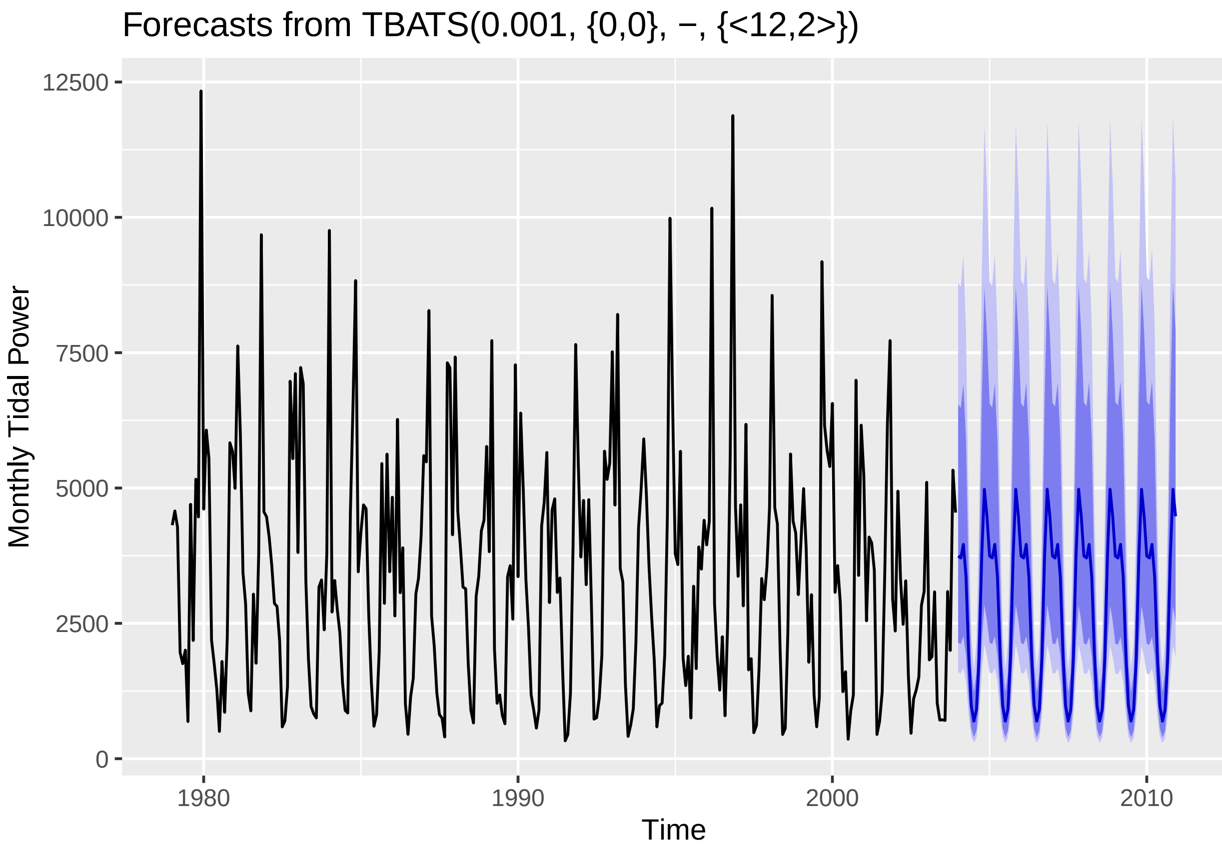
Model 2: ARIMA + Fourier terms

```
#Model 2: Arima + Fourier
k <- 6 #Value that gives lowest RMSE - need to justify
florida_arima_fit <- auto.arima(florida_ts_train,
                                seasonal=FALSE,
                                lambda=0,
                                xreg=fourier(florida_ts_train,
                                             K = k))

florida_arima_fore <- forecast(florida_arima_fit,
                              xreg=fourier(florida_ts_train,
                                           K = k,
                                           h = fore_h),
                              h = fore_h)
autoplot(florida_arima_fore) + ggtitle("ARIMA + Fourier Forecast - Florida")
```

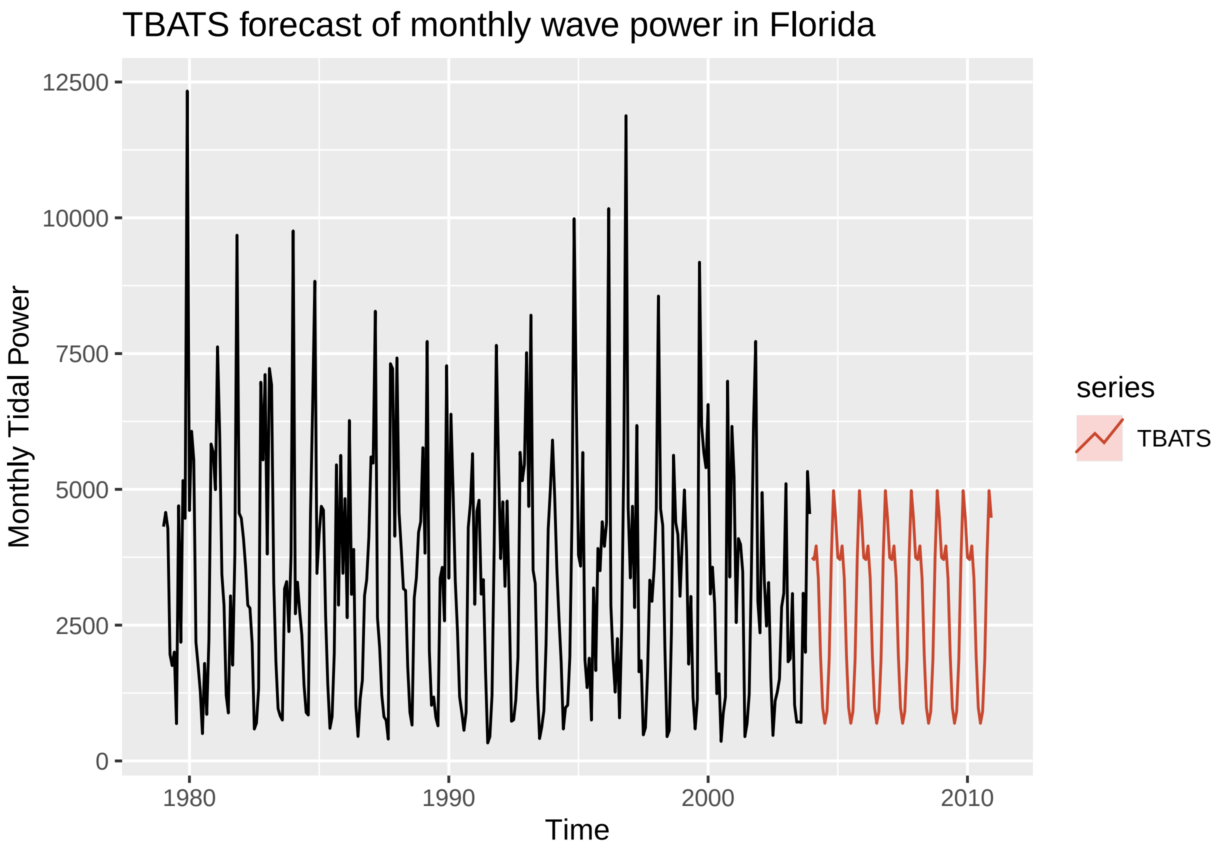


Model 3 TBATs

```
#Model 3: TBATS
florida_train_tbats_fit <- tbats(florida_ts_train)

tbats_forecast <- forecast(florida_train_tbats_fit, h=fore_h)

#Plot foresting results
autoplot(tbats_forecast) +
  ylab("Monthly Tidal Power")
```




```
#Plot model + observed data
autoplot(florida_ts_train) +
  autolayer(tbats_forecast, series="TBATS",PI=FALSE)+
  ylab("Monthly Tidal Power") +
  ggtitle("TBATS forecast of monthly wave power in Florida")
```




```r
#summarizing the scores/results from the above models
ETS_scores <- accuracy(florida_stlf_ets$mean,florida_ts_test)
ARIMAF_scores <- accuracy(florida_arima_fore$mean,florida_ts_test)
TBATS_scores <- accuracy(tbats_forecast$mean,florida_ts_test)

scores <- as.data.frame(
  rbind(ETS_scores, ARIMAF_scores, TBATS_scores)
  )
row.names(scores) <- c("STL+ETS", "ARIMA+Fourier", "TBAT")

#choose model with lowest RMSE
best_model_index <- which.min(scores[,"RMSE"])
cat("The best model by RMSE is:", row.names(scores[best_model_index,]))
```

```
## The best model by RMSE is: STL+ETS
```

```r
kbl(scores,
      caption = "Forecast Accuracy for Monthly Wave Power - Florida",
      digits = array(5,ncol(scores))) %>%
  kable_styling(full_width = FALSE, position = "center", latex_options = "hold_position") %>%
  #highlight model with lowest RMSE
  kable_styling(latex_options="striped", stripe_index = which.min(scores[,"RMSE"]))
```

Table 3: Forecast Accuracy for Monthly Wave Power - Florida

	ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's U
STL+ETS	260.0615	1884.797	1242.601	-22.94636	48.47699	0.04601	0.78809
ARIMA+Fourier	377.5698	2019.667	1254.629	-13.78776	42.55154	0.03404	0.85821
TBAT	644.5316	2081.479	1271.549	-4.49300	39.17815	0.02116	0.86973