Arctic Sea Ice Extent

Tim Creed

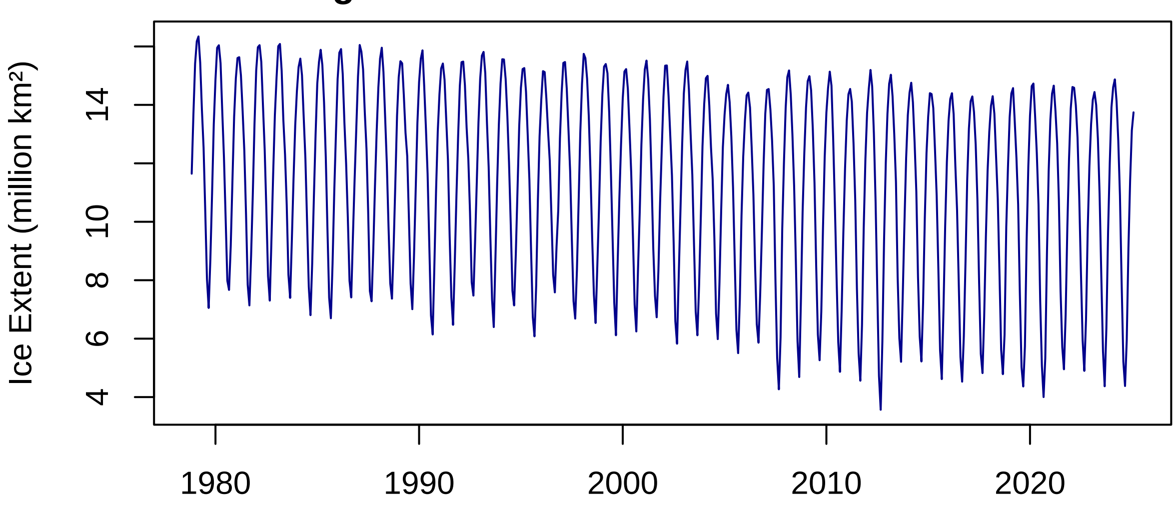
2025-06-13

Introduction

Due to climate change, Arctic and Antarctic sea ice has steadily declined over the past few decades. Sea ice plays a crucial role in regulating the Earth's temperature by reflecting sunlight back into space (Krishna, 2023). With the sea ice melting more and more over time, this creates a feedback loop that accelerates global warming. Sea ice also acts as a habitat for many marine species, and the melting of the ice disrupts the marine life ecosystem (Qi et al., 2024). Lastly, the loss of sea ice affects weather patterns and ocean currents, which may lead to harmful changes in climate patterns across the world (Nihoul & Kostianoy, 2009).

Data Description The Sea Ice Index, maintained by the National Snow and Ice Data Center (NSIDC), provides data on Arctic and Antarctic sea ice extent and area from November 1978 to the present (Fetterer et al., 2017). Sea ice extent is defined as the total area where at least 15% of the surface is covered by sea ice, offering a standardized way to track changes over time. Sea ice extent provides a more consistent measure of ice coverage that is less sensitive to short-term fluctuations when compared to sea ice area. We use the monthly data, as recommended by NSIDC, to reduce errors from short-term weather fluctuations present in daily data. The dataset includes Ice Extent and Ice Area for both hemispheres, with our analysis focusing on Arctic sea ice extent as the primary parameter. By analyzing these trends, we aim to quantify ice loss and project future changes in sea ice cover.

Figure 1. Arctic Sea Ice Extent Over Time



Methods

There were two missing values in the data set. Kalman filtering was used to impute the missing values, preserving the time series structure and temporal patterns.

Next, the sea ice extent values were plotted vs time (Figure 1). The data shows a clear pattern of annual seasonality and an overall decreasing trend (Figure 2-3). Linear models fit against month and year were used to isolate seasonality and trend.

After removing seasonality and trend, the Augmented Dickey-Fuller Test was used to confirm that the residual series is stationary (p-value < 0.01).

After analyzing the ACF and PACF plots of the residual series (Figure 5), different combinations of parameters for ARMA and SARIMA models were tested. The SARIMA model with parameters (1,0,1)(2,0,0)[12] showed the best fit when comparing AIC and Log Likelihood values.

Model diagnostics show that the SARIMA(1,0,1)(2,0,0)[12] model provides an adequate fit to the data. The ACF and PACF of the residuals and Ljung-Box test show no significant autocorrelation (Figure 6). The residuals appear normally distributed based on the Q-Q plot (Figure 7). With the model assumptions met, it can be used for forecasting. This model was used to create a 10 year forecast with 95% confidence intervals included (Figure 8).

We also applied a Holt-Winters exponential smoothing model with multiplicative seasonality to forecast Arctic sea ice extent, optimizing initial parameters and using a 95% confidence level for a 10 year projection (Figure 9).

Figure 2. Seasonality

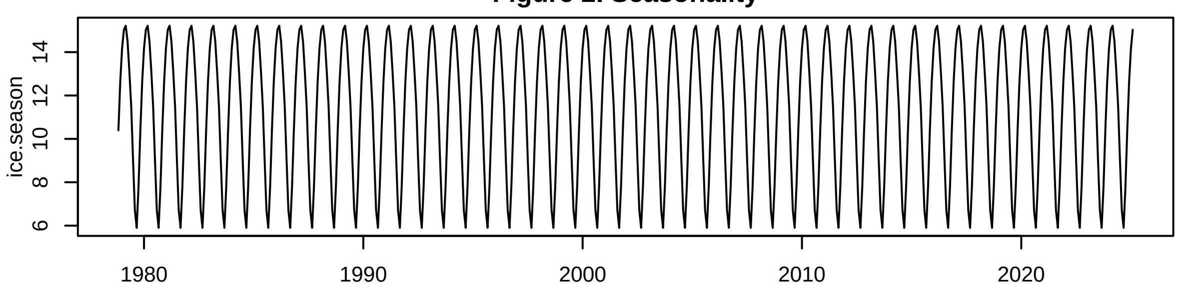


Figure 3. Trend

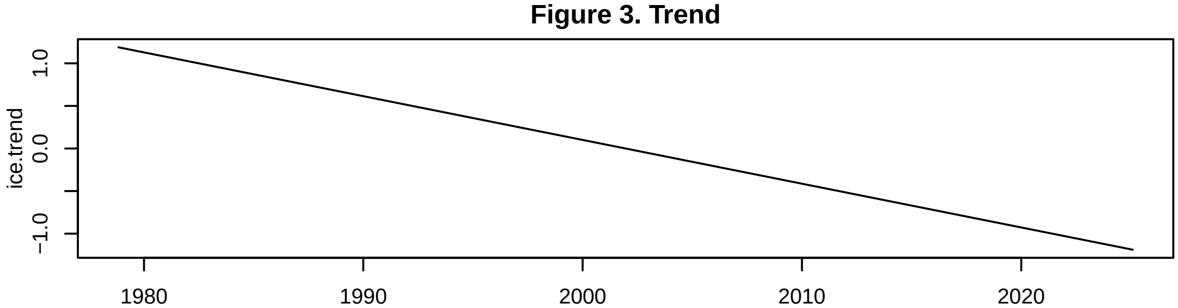


Figure 4. Residual Series

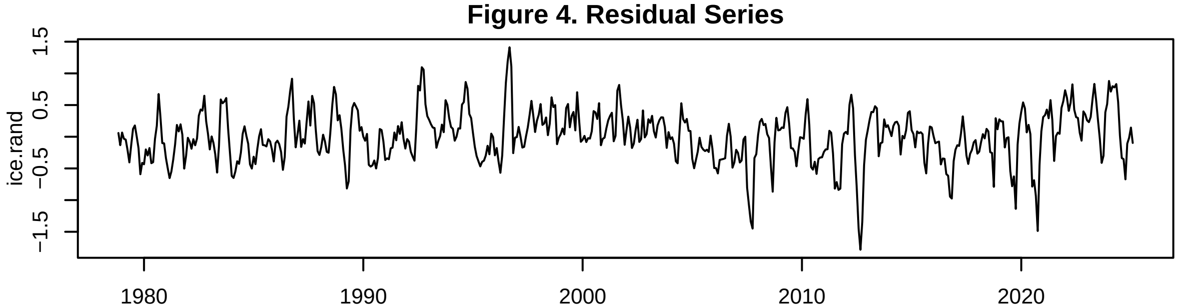


Figure 5. Residual series plots

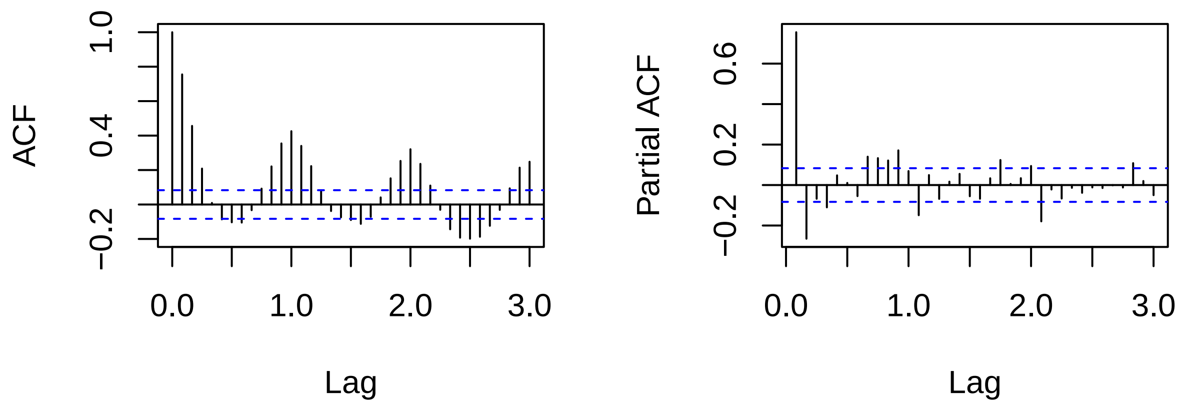


Figure 6. SARIMA(1,0,1)(2,0,0)[12] - Residual Diagnostic Plots

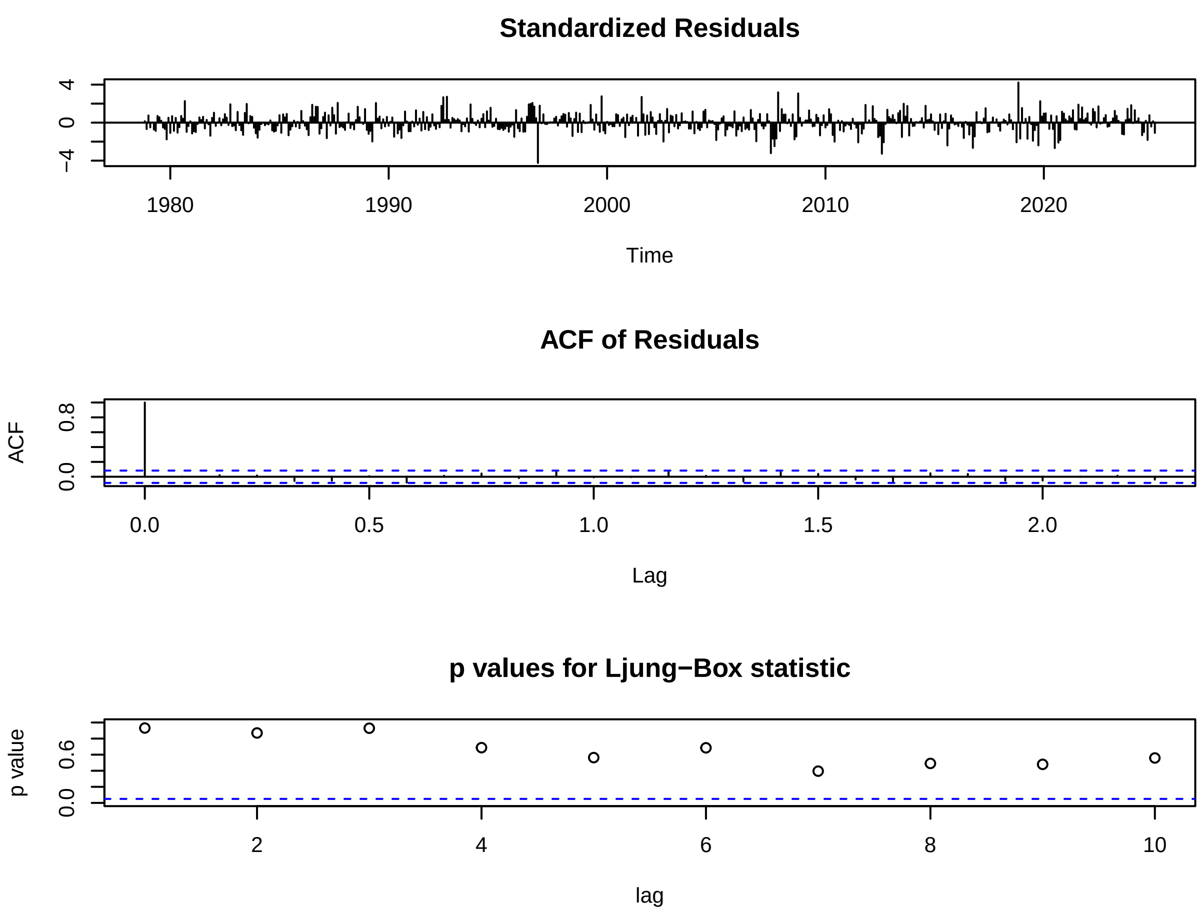
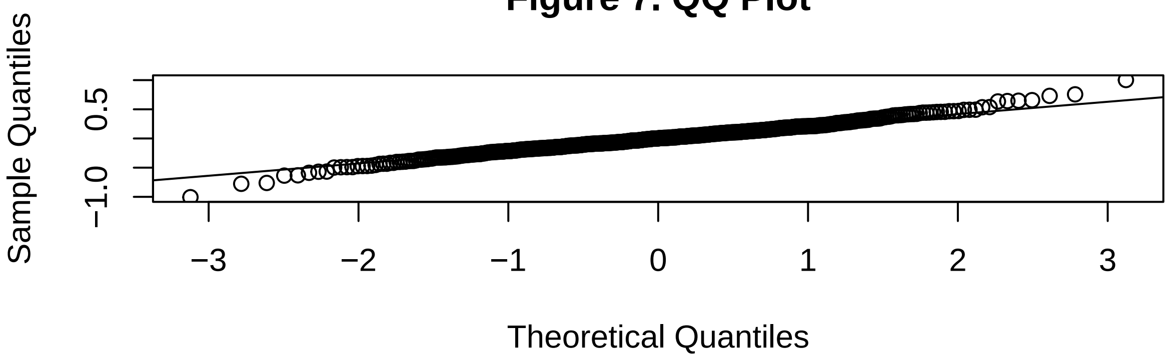


Figure 7. QQ Plot



Results

Based on a linear model, sea ice extent is decreasing annually by approximately 51400 km². With 95% confidence, the true annual decrease is between 48900 and 53900 million km².

The best fitting model was a SARIMA model with parameters (1,0,1)(2,0,0)[12], and has a Log-likelihood of 9.39 and AIC of -8.79.

The current Arctic Sea Ice extent for February 2025 is 13.745 million km². Using the SARIMA forecast, we predict that in 10 years, there will be a loss of 412000 km² sea ice. With 95% confidence, the sea ice loss will be between 1526000 and -701000 km².

Using the SARIMA forecast, we predict that in 10 years, there will be a loss of 219000 km² sea ice. With 95% confidence, the sea ice loss will be between 2096000 and -1657000 km².

Figure 8. Arctic Sea Ice Extent with SARIMA Forecast

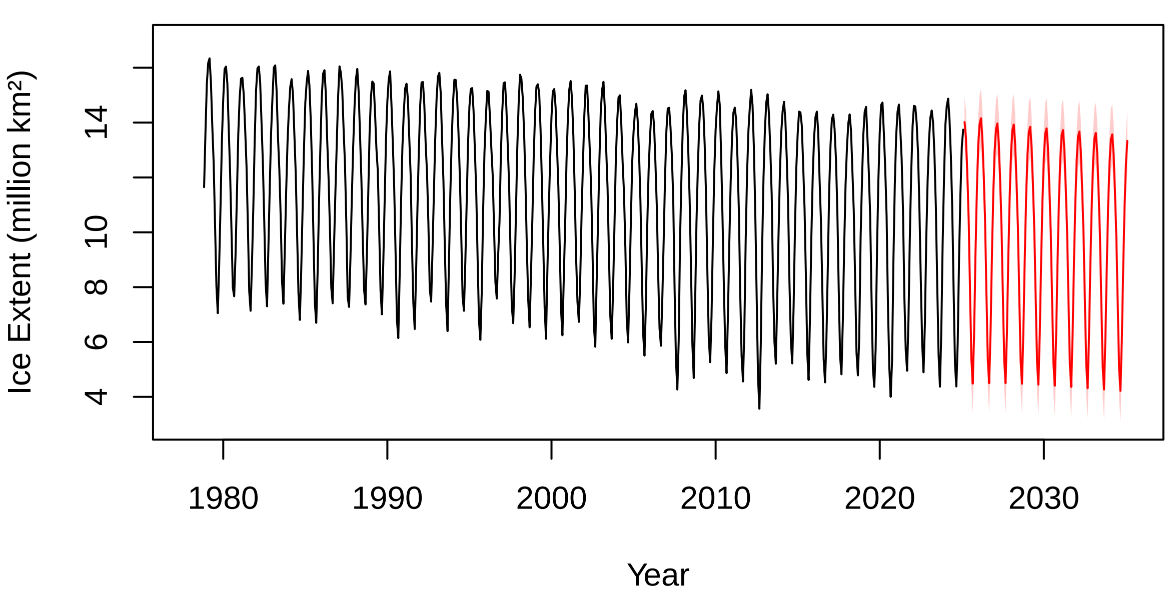
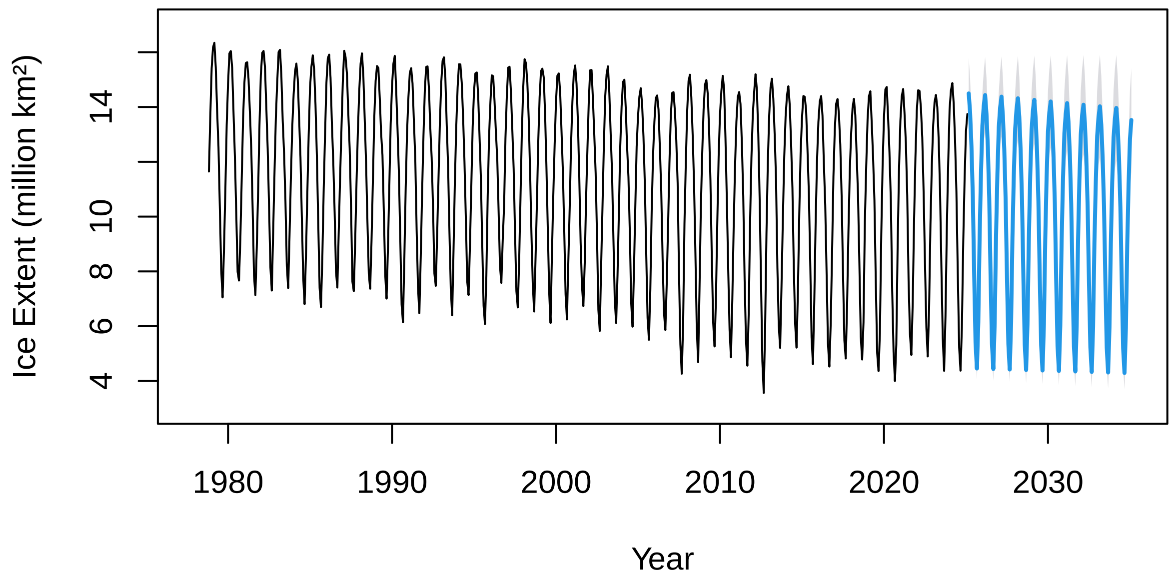


Figure 9. Arctic Sea Ice Extent with Holt−Winters Forecast



Discussion

The goal of this analysis was to forecast the Arctic sea ice data. This was accomplished through imputing missing data through Kalman Filtering, isolating the seasonality and trend, and forecasting the data using a SARIMA model and Holt-Winters exponential smoothing.

The initial time series plot (Figure 1) shows how Arctic sea ice has changed from 1978 to 2024. We can see strong annual seasonality with sea ice reaching maximums in the winter and minimums in the summer. There is also a long-term trend showing that sea ice is gradually decreasing over time.

The mean annual ice extent for the first five full years of the study, 1979-1983, is 12.33 million km^2. The mean annual ice extent for 2019-2023 is 10.47 million km^2. While ice coverage still follows seasonal patterns, we are losing more and more of it over time. The annual ice extent maximums do not recover to past levels. We estimate the average annual decline in Arctic sea ice extent to be over 50000 km^2.

This is a big concern because less ice means more heat absorption in the Arctic, which can speed up warming and affect weather, ocean currents, and wildlife. If this trend continues, we could see even more ice loss, leading to bigger climate impacts worldwide.

References

Fetterer, F., Knowles, K., Meier, W. N., Savoie, M. & Windnagel, A. K. (2017). Sea Ice Index. (G02135, Version 3). [Data Set]. Boulder, Colorado USA. National Snow and Ice Data Center. https://doi.org/10.7265/N5K072F8. [describe subset used if applicable]. Date Accessed 03-14-2025.

Krishna, K. M. (2023). Spatial and Temporal Variations of Arctic Sea Ice during 1992-2021. https://doi.org/10.19080/ofoaj.2023.16.555935

Nihoul, J., & Kostianoy, A. (2009). Influence of Climate Change on the Changing Arctic and Sub-Arctic Conditions. Springer, published in cooperation with NATO Public Diplomacy Division. https://doi.org/10.1007/978-1-4020-9460-6

Qi, Q., Hao, Q., Yang, G., Cao, S., Kang, J., Hu, J., Zheng, M., Chen, M., He, J., & Chen, M. (2024). Diverse impacts of sea ice and ice shelf melting on phytoplankton communities in the Cosmonaut Sea, East Antarctica. Environmental Research Letters. https://doi.org/10.1088/1748-9326/ad975e

Appendix

```
## ----setup, include=FALSE----
knitr::opts_chunk$set(echo = TRUE)
library(imputeTS)
library(forecast)
library(tseries)
library(tidyverse)


## ----echo = FALSE, warning=FALSE, message=FALSE, fig.height = 3----
# import first sheet NH-Extent
# you'll need to change the path to your file path
ice <- readxl::read_xlsx("Sea_Ice_Index_Monthly_Data_by_Year_G02135_v3.0.xlsx", sheet = 1)

# remove last two columns
ice <- ice[,-c(ncol(ice)-1, ncol(ice))]

# put all data into one numerical vector row-wise
ice_values <- c(t(ice[,-1]))

# eliminate leading and trailing na values (anything before Nov 1978 or after Feb 2025)
ice_values <- ice_values[11:566]

# best method I could find for imputing the two missing values
ice_values <- na.kalman(ice_values)
#is.na(ice_values)

ice.ts <- ts(ice_values, start = c(1978, 11), deltat = 1/12)

par(mar = c(2.5, 4, 2, 2))
plot(ice.ts, main = "Figure 1. Arctic Sea Ice Extent Over Time", ylab = "Ice Extent (million km²)", col

ice_means <- ice |>
  rowwise() |>
  mutate(mean_extent = mean(c_across(-1), na.rm = TRUE)) |>
  ungroup()

first_five_mean <- mean(ice_means$mean_extent[2:6], na.rm = TRUE)
last_five_mean <- mean(ice_means$mean_extent[(nrow(ice)-5):(nrow(ice)-1)], na.rm = TRUE)


## ----echo = FALSE, out.width="100%", fig.align="center", fig.height=5----
ice.time <- time(ice.ts)
```

```r
ice.month <- factor(cycle(ice.ts))

# seasonal model
fit.season <- lm(ice.ts ~ ice.month)
ice.season <- ts(fit.season$fitted.values, start = c(1978, 11), deltat = 1/12)

# remove seasonality
ice.deseasonalized <- ice.ts - ice.season

# trend on deseasonalized data
ice.lm <- lm(ice.deseasonalized ~ ice.time)
ice.trend <- ts(ice.lm$fitted.values, start = c(1978, 11), deltat = 1/12)
#summary(ice.lm)
#confint(ice.lm)

# remove trend
ice.rand <- ts(ice.deseasonalized - ice.trend, start = c(1978, 11), deltat = 1/12)

# plots
par(mfrow = c(3, 1), mar = c(2, 4, 2, 2), oma = c(0, 0, 0, 0), mgp = c(2, 1, 0))
plot(ice.season, main = "Figure 2. Seasonality")
plot(ice.trend, main = "Figure 3. Trend")
plot(ice.rand, main = "Figure 4. Residual Series")

#adf.test(ice.rand)
# Dickey-Fuller = -5.7043, Lag order = 8, p-value = 0.01, series is stationary


## ----echo = FALSE, fig.height=3----
par(mfrow = c(1, 2))
acf(ice.rand, main = "Figure 5. Residual series plots",lag.max=36)
pacf(ice.rand, main = "", lag.max=36)

fit.arma201 <- arima(ice.rand, order = c(p = 2, d = 0, q = 1), method = "ML", include.mean = F)
#sigma^2 = 0.06548:  log likelihood = -30.12 AIC=68.24   AICc=68.32   BIC=85.53

fit.arma101 <- arima(ice.rand, order = c(p = 1, d = 0, q = 1), method = "ML", include.mean = F)
#sigma^2 estimated as 0.06639:  log likelihood = -35.42,  aic = 76.85

fit.arma202 <- arima(ice.rand, order = c(p = 2, d = 0, q = 2), method = "ML", include.mean = F)
# sigma^2 estimated as 0.06511:  log likelihood = -30.04,  aic = 70.08

fit.101.200 <- arima(ice.rand, order = c(1,0,1), seasonal = list(order = c(2,0,0), period = 12))
# sigma^2 = 0.05665:  log likelihood = 9.39, aic = -8.79

fit.101.100 <- arima(ice.rand, order = c(1,0,1), seasonal = list(order = c(1,0,0), period = 12))
#sigma^2 estimated as 0.05952:  log likelihood = -5.69,  aic = 21.38

fit.201.200 <- arima(ice.rand, order = c(2,0,1), seasonal = list(order = c(2,0,0), period = 12))
#sigma^2 estimated as 0.05621:  log likelihood = 9.61,  aic = -5.22

#auto arima chose (1,0,1)(2,0,0)[12]
best_model <- auto.arima(ice.rand, seasonal = TRUE)
```

```r
#summary(best_model)


## ----echo = FALSE, fig.height=4----
#Using fit.101.200
#checkresiduals(fit.101.200)

# par(mfrow = c(1, 2))
# acf(fit.101.200$residuals, lag.max = 36, main = "ACF for model residuals")
# pacf(fit.101.200$residuals, lag.max = 36, main = "PACF for model residuals")


## ----echo = FALSE, fig.height=5----
tsdiag(fit.101.200)


## ----echo = FALSE, fig.height=2.5----
qqnorm(fit.101.200$residuals, main = "Figure 7. QQ Plot")
qqline(fit.101.200$residuals)


## ----echo = FALSE, fig.height= 4----
rand.pred <- predict(fit.101.200, n.ahead = 120)
rand.fit <- rand.pred$pred
rand.se <- rand.pred$se


# prediction time starts at MArch 2025
pred.time <- seq(2025 + (3 - 1) / 12, by = 1/12, length = 120)

# trend predictions
trend.pred <- predict(ice.lm, newdata = data.frame(ice.time = pred.time), interval = "prediction")
trend.fit <- trend.pred[, "fit"]
trend.lower <- trend.pred[, "lwr"]
trend.upper <- trend.pred[, "upr"]


# seasonal predictions, two years starting in march
season.pred <- fit.season$fitted[5:124]

# final prediction time series
pred <- ts(rand.fit + trend.fit + season.pred, start = c(2025, 3), deltat = 1/12)

# combined standard error
total.se <- sqrt(rand.se^2 + (trend.pred[, "upr"] - trend.pred[, "fit"])^2 / (1.96^2))

#  final confidence intervals
z_critical <- 1.96
pred.lower <- ts(rand.fit + trend.fit + season.pred - z_critical * total.se, start = c(2025, 3), deltat
pred.upper <- ts(rand.fit + trend.fit + season.pred + z_critical * total.se, start = c(2025, 3), deltat
```

```
# Plot observed data with forecast
plot(ice.ts, xlim = c(1978, 2035), xlab = "Year", ylab = "Ice Extent (million km²)",
     main = "Figure 8. Arctic Sea Ice Extent with SARIMA Forecast", ylim = c(3, 17))
lines(pred, col = "red")
polygon(c(time(pred), rev(time(pred))),
        c(pred.upper, rev(pred.lower)),
        col = rgb(1, 0, 0, 0.2), border = NA)

# Add upper and lower bounds as dashed lines
#lines(pred.upper, col = "blue", lty = 2)
#lines(pred.lower, col = "blue", lty = 2)


## ----echo = FALSE, fig.height= 4----
ice.hw <- hw(ice.ts, initial = "optimal", level = 95, seasonal = "multiplicative", h = 120)
plot(ice.hw, xlab = "Year", ylab = "Ice Extent (million km²)", xlim = c(1978, 2035), ylim = c(3, 17),
     main = "Figure 9. Arctic Sea Ice Extent with Holt-Winters Forecast")


## ----echo = TRUE, eval = FALSE, message=FALSE, warning=FALSE----
# ## ----setup, include=FALSE--------------------------------------------------------------------
# knitr::opts_chunk$set(echo = TRUE)
# library(imputeTS)
# library(forecast)
# library(tseries)
#
#
# ## ----echo = FALSE, warning=FALSE, message=FALSE, fig.height = 3-----------------------------------
# # import first sheet NH-Extent
# # you'll need to change the path to your file path
# ice <- readxl::read_xlsx("/Users/jcorbett/Documents/ST566/Sea_Ice_Index_Monthly_Data_by_Year_G02135_v
#
# # remove last two columns
# ice <- ice[,-c(ncol(ice)-1, ncol(ice))]
#
# # put all data into one numerical vector row-wise
# ice_values <- c(t(ice[,-1]))
#
# # eliminate leading and trailing na values (anything before Nov 1978 or after Feb 2025)
# ice_values <- ice_values[11:566]
#
# # best method I could find for imputing the two missing values
# ice_values <- na.kalman(ice_values)
# #is.na(ice_values)
#
# ice.ts <- ts(ice_values, start = c(1978, 11), deltat = 1/12)
#
# par(mar = c(2.5, 4, 2, 2))
# plot(ice.ts, main = "Figure 1. Arctic Sea Ice Extent Over Time", ylab = "Ice Extent (million km²)", c
#
# ice_means <- ice |>
#   rowwise() |>
#   mutate(mean_extent = mean(c_across(-1), na.rm = TRUE)) |>
```

```
#   ungroup()
#
# first_five_mean <- mean(ice_means$mean_extent[2:6], na.rm = TRUE)
# last_five_mean <- mean(ice_means$mean_extent[(nrow(ice)-5):(nrow(ice)-1)], na.rm = TRUE)
#
# ## ----echo = FALSE, out.width="100%", fig.align="center", fig.height=5----------------------------
# ice.time <- time(ice.ts)
# ice.month <- factor(cycle(ice.ts))
#
# # seasonal model
# fit.season <- lm(ice.ts ~ ice.month)
# ice.season <- ts(fit.season$fitted.values, start = c(1978, 11), deltat = 1/12)
#
# # remove seasonality
# ice.deseasonalized <- ice.ts - ice.season
#
# # trend on deseasonalized data
# ice.lm <- lm(ice.deseasonalized ~ ice.time)
# ice.trend <- ts(ice.lm$fitted.values, start = c(1978, 11), deltat = 1/12)
# #summary(ice.lm)
# #confint(ice.lm)
#
# # remove trend
# ice.rand <- ts(ice.deseasonalized - ice.trend, start = c(1978, 11), deltat = 1/12)
#
# # plots
# par(mfrow = c(3, 1), mar = c(2, 4, 2, 2), oma = c(0, 0, 0, 0), mgp = c(2, 1, 0))
# plot(ice.season, main = "Figure 2. Seasonality")
# plot(ice.trend, main = "Figure 3. Trend")
# plot(ice.rand, main = "Figure 4. Residual Series")
#
# #adf.test(ice.rand)
# # Dickey-Fuller = -5.7043, Lag order = 8, p-value = 0.01, series is stationary
#
# ## ----echo = FALSE, fig.height=3------------------------------------------------------------------
# par(mfrow = c(1, 2))
# acf(ice.rand, main = "Figure 5. Residual series plots",lag.max=36)
# pacf(ice.rand, main = "", lag.max=36)
#
# fit.arma201 <- arima(ice.rand, order = c(p = 2, d = 0, q = 1), method = "ML", include.mean = F)
# #sigma^2 = 0.06548:  log likelihood = -30.12 AIC=68.24   AICc=68.32   BIC=85.53
#
# fit.arma101 <- arima(ice.rand, order = c(p = 1, d = 0, q = 1), method = "ML", include.mean = F)
# #sigma^2 estimated as 0.06639:  log likelihood = -35.42,  aic = 76.85
#
# fit.arma202 <- arima(ice.rand, order = c(p = 2, d = 0, q = 2), method = "ML", include.mean = F)
# # sigma^2 estimated as 0.06511:  log likelihood = -30.04,  aic = 70.08
#
# fit.101.200 <- arima(ice.rand, order = c(1,0,1), seasonal = list(order = c(2,0,0), period = 12))
# # sigma^2 = 0.05665:  log likelihood = 9.39, aic = -8.79
#
# fit.101.100 <- arima(ice.rand, order = c(1,0,1), seasonal = list(order = c(1,0,0), period = 12))
# #sigma^2 estimated as 0.05952:  log likelihood = -5.69,  aic = 21.38
```

```
#
# fit.201.200 <- arima(ice.rand, order = c(2,0,1), seasonal = list(order = c(2,0,0), period = 12))
# #sigma^2 estimated as 0.05621:  log likelihood = 9.61,  aic = -5.22
#
# #auto arima chose (1,0,1)(2,0,0)[12]
# best_model <- auto.arima(ice.rand, seasonal = TRUE)
# #summary(best_model)
#
#
# ## ----echo = FALSE, fig.height=4-----------------------------------------------------------------
# #Using fit.101.200
# #checkresiduals(fit.101.200)
#
# # par(mfrow = c(1, 2))
# # acf(fit.101.200$residuals, lag.max = 36, main = "ACF for model residuals")
# # pacf(fit.101.200$residuals, lag.max = 36, main = "PACF for model residuals")
#
#
# ## ----echo = FALSE, fig.height=5-----------------------------------------------------------------
# tsdiag(fit.101.200)
#
#
# ## ----echo = FALSE, fig.height=2.5---------------------------------------------------------------
# qqnorm(fit.101.200$residuals, main = "Figure 6. QQ Plot")
# qqline(fit.101.200$residuals)
#
#
# ## ----echo = FALSE, fig.height= 4----------------------------------------------------------------
# rand.pred <- predict(fit.101.200, n.ahead = 120)
# rand.fit <- rand.pred$pred
# rand.se <- rand.pred$se
#
#
# # prediction time starts at MArch 2025
# pred.time <- seq(2025 + (3 - 1) / 12, by = 1/12, length = 120)
#
# # trend predictions
# trend.pred <- predict(ice.lm, newdata = data.frame(ice.time = pred.time), interval = "prediction")
# trend.fit <- trend.pred[, "fit"]
# trend.lower <- trend.pred[, "lwr"]
# trend.upper <- trend.pred[, "upr"]
#
#
# # seasonal predictions, two years starting in march
# season.pred <- fit.season$fitted[5:124]
#
# # final prediction time series
# pred <- ts(rand.fit + trend.fit + season.pred, start = c(2025, 3), deltat = 1/12)
#
# # combined standard error
# total.se <- sqrt(rand.se^2 + (trend.pred[, "upr"] - trend.pred[, "fit"])^2 / (1.96^2))
#
# #  final confidence intervals
```

```r
# z_critical <- 1.96
# pred.lower <- ts(rand.fit + trend.fit + season.pred - z_critical * total.se, start = c(2025, 3), delt
# pred.upper <- ts(rand.fit + trend.fit + season.pred + z_critical * total.se, start = c(2025, 3), delt
#
#
#
# # Plot observed data with forecast
# plot(ice.ts, xlim = c(1978, 2035), xlab = "Year", ylab = "Ice Extent (million km²)",
#      main = "Figure 7. Arctic Sea Ice Extent with SARIMA Forecast", ylim = c(3, 17))
# lines(pred, col = "red")
# polygon(c(time(pred), rev(time(pred))),
#         c(pred.upper, rev(pred.lower)),
#         col = rgb(1, 0, 0, 0.2), border = NA)
#
# # Add upper and lower bounds as dashed lines
# #lines(pred.upper, col = "blue", lty = 2)
# #lines(pred.lower, col = "blue", lty = 2)
#
#
# ## ----echo = FALSE, fig.height= 4------------------------------------------------------------------
# ice.hw <- hw(ice.ts, initial = "optimal", level = 95, seasonal = "multiplicative", h = 120)
# plot(ice.hw, xlab = "Year", ylab = "Ice Extent (million km²)", xlim = c(1978, 2035), ylim = c(3, 17),
#      main = "Figure 8. Arctic Sea Ice Extent with Holt-Winters Forecast")
#
#
#



knitr::purl("arctic_ice_17Mar2025.Rmd", "extracted_code2.R")
```