Introduction

Since February 2022, the unforeseen Russia-Ukraine conflict has created significant disruptions in the global grain market, sparking widespread concerns over food insecurity issues. As the world's second-largest oat exporter in 2021, Canada inevitably faced the challenges of navigating these market uncertainties. Notably, British Columbia, renowned for its fertile lands, plays a crucial role as one of Canada's major oat producers.

Our project aims to provide a detailed examination and a comprehensive time series analysis of oat producer delivery data in British Columbia. Preliminary examination of the data suggests that an ARIMA model is optimal for handling the trend and seasonality presented in the time series. This allows us to provide accurate forecasts for future oat production in the province. The ultimate objective is to promote the formulation of targeted policies that can effectively support local farmers in the wake of a volatile global market.

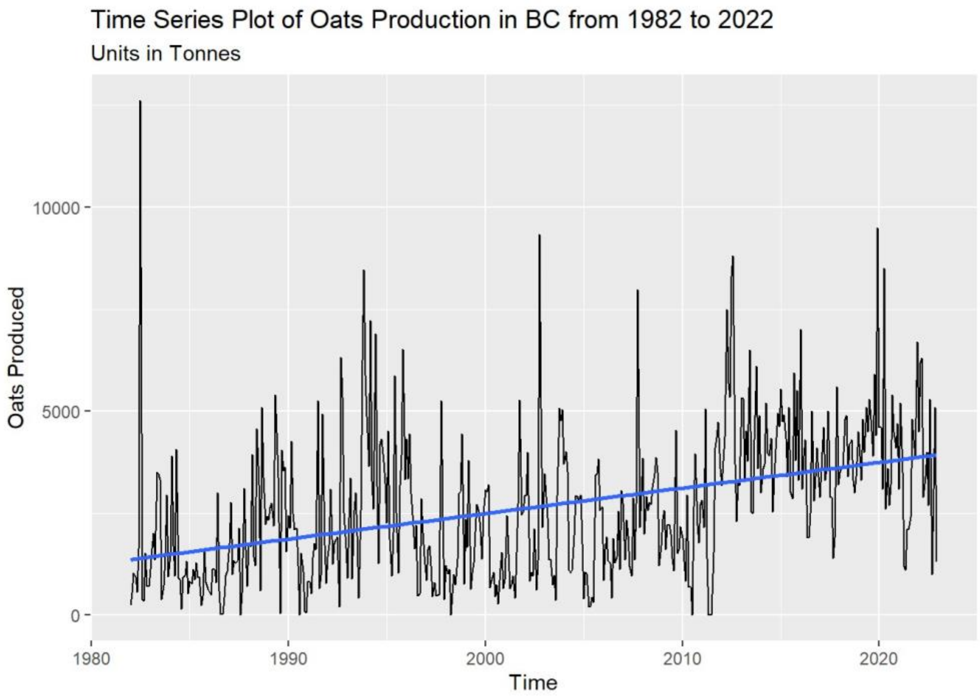
Data Description

We obtained our dataset from the official website of Canada's national statistical agency, Statistics Canada. The dataset consists of historical data collected monthly on the producer deliveries of major grains, including wheat, durum, oats, barley, rye, flaxseed, and Canola. We restricted our attention to oats production in British Columbia from January 1982 to December 2022, with monthly production of oats recorded on the first day of each corresponding month. In particular, We are interested in making informed predictions on future oats production (in tonnes) in British Columbia. Observations with reference dates outside of this time interval were dropped to ensure proper handling of potential seasonality in the dataset.

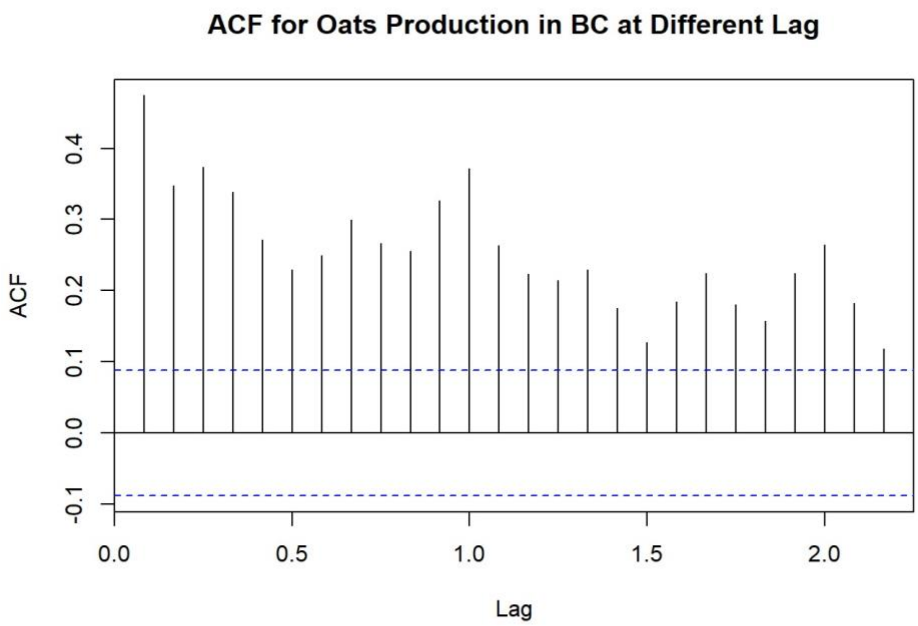
Model Identification

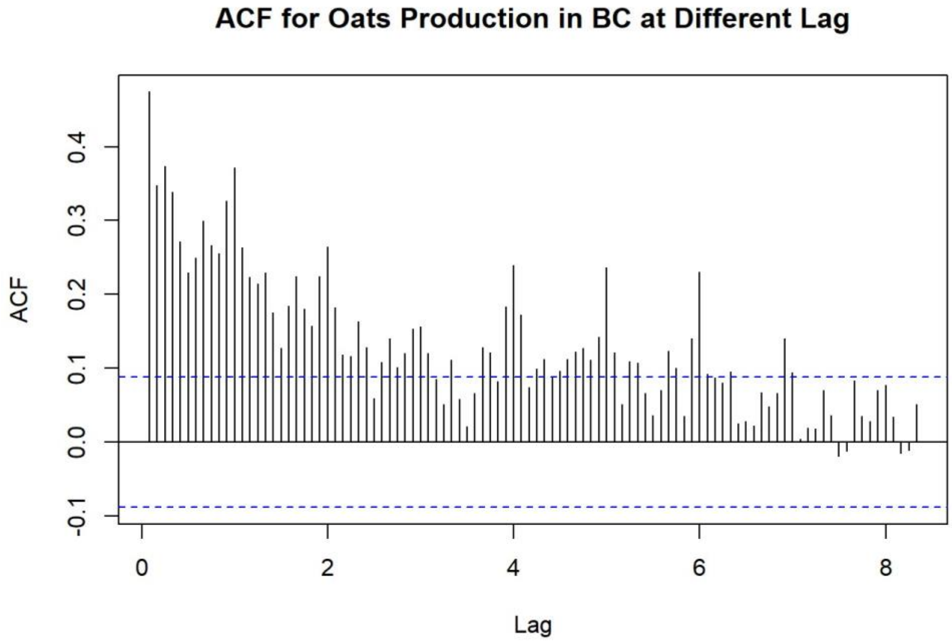
To make an accurate forecast in terms of BC's future oats production, we first inspect the time series data by plotting historical quantities produced for oats v.s. time to check for potential trends, patterns, and seasonality that may comprise our selection of appropriate orders for an ARIMA model. Our time series plot illustrates the total oat production for each month in British Columbia, Canada, spanning from January 1982 to December 2022. The independent variable is the time, while the dependent variable is the production quantity of oats measured in tonnes. There appears to be an upward trend in the monthly production of oats from 1982 onwards, suggesting a long-term growth of oats produced in BC. This implies that the time series might be non-stationary as reflected by a changing mean. By mere inspection, we did not see a clear sign of seasonality in the data as no distinguishable periodic patterns are recurring over time. However, the dataset claimed not to be seasonally adjusted, reflecting the periodic nature of the data. To avoid producing biased predictions, we then plotted the autocorrelation function (ACF) and the partial autocorrelation function (PACF) to further investigate the dataset.

Time Series Plot



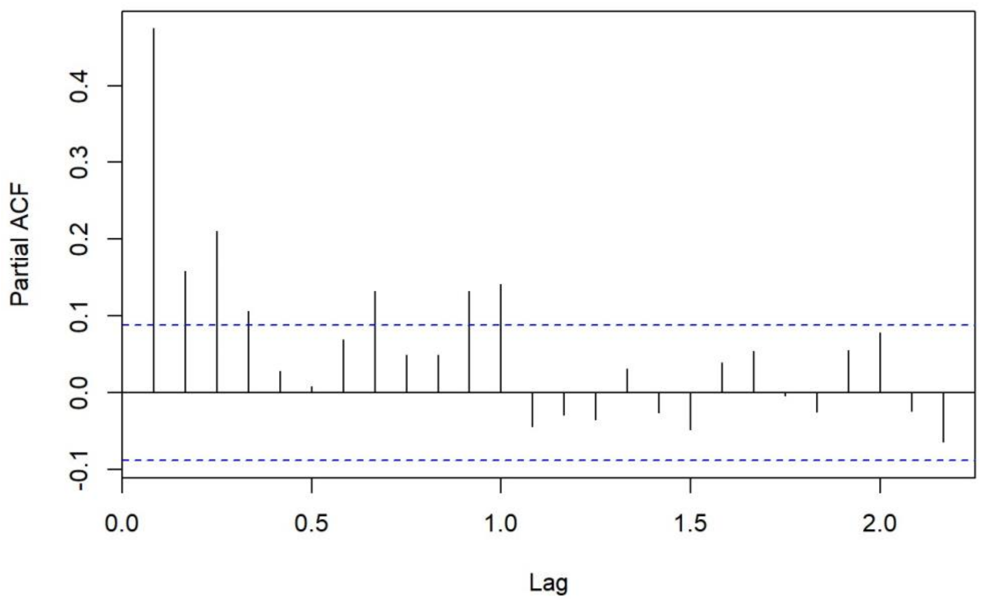
Plots of ACF and PACF



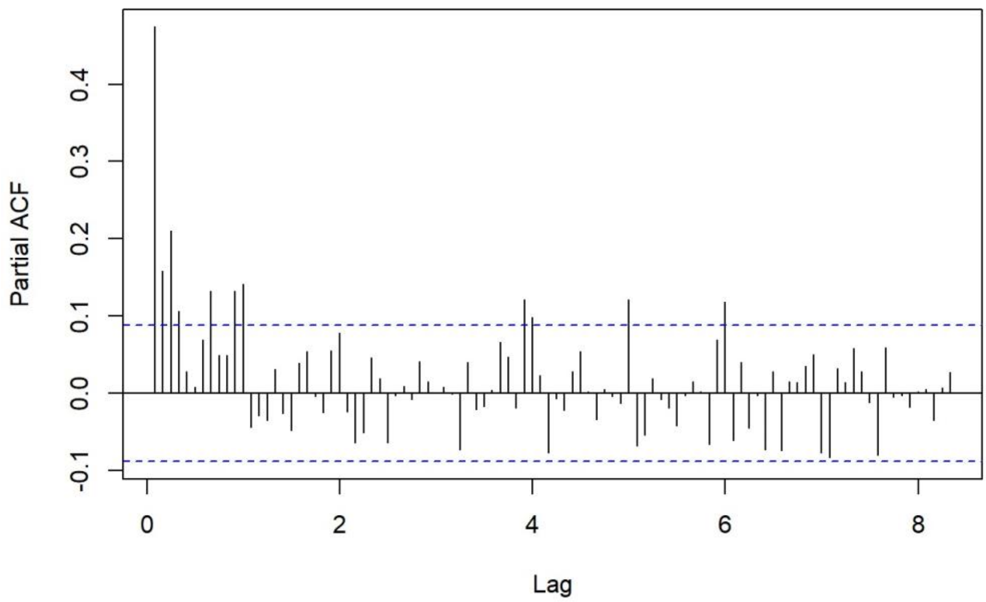


Since the dataset contains a relatively large number of observations, we decided to plot the ACF of the time series with more lags. Hence, we set the maximum range of lag to 100 in the second ACF plot. The x-axises of both plots had been altered due to seasonality, indicating that a lag of 12 corresponds to a one-year cycle.

PACF for Oats Production in BC at Different Lag



PACF for Oats Production in BC at Different Lag



Similarly, we set the maximum order of k equivalent to 100 in the PACF plots to capture seasonal patterns.

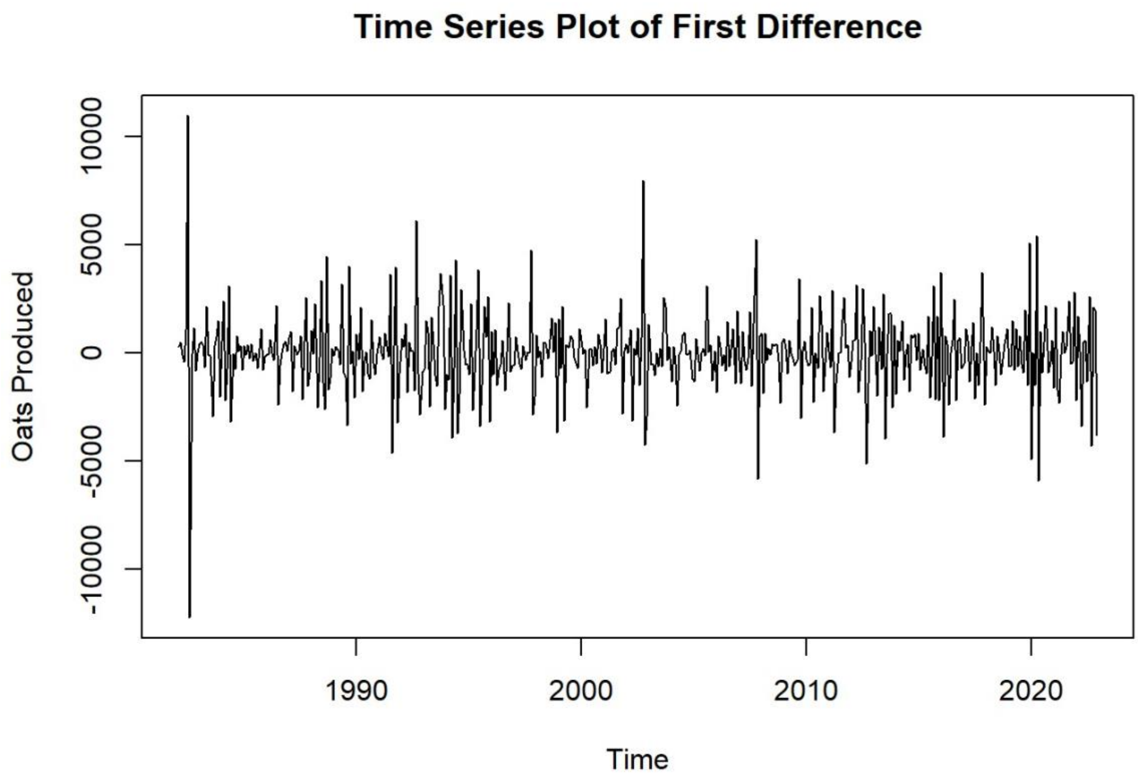
Unit Root Test Output - Augmented Dickey-Fuller Method

```
        Augmented Dickey-Fuller Test

data:  oat_dt_tsa
Dickey-Fuller = -2.2637, Lag order = 50, p-value = 0.4666
alternative hypothesis: stationary
```

Collectively, it is evident from the ACF and PACF plots that the time series is not stationary. The AFC plots show slow decay of autocorrelation value as lag increases, indicating the persistent dependence of observations over time. We can also see multiple spikes at various lags. Moreover, there appears to be a seasonal pattern of the middle of every year being the month with the least oat production, signifying the presence of non-stationarity due to seasonality in the data. The PACF plots support our findings as both are slowly tailing, meaning partial autocorrelation values decrease gradually. Furthermore, we used the acf.test() command from the tseries package to confirm the existence of the non-stationarity of our data. The output reports that, with more lags (k = 50), the t-statistic is -2.2637 with a p-value of 0.4666, exceeding the default level of significance level of 0.05. We thus conclude that our time series is non-stationary.

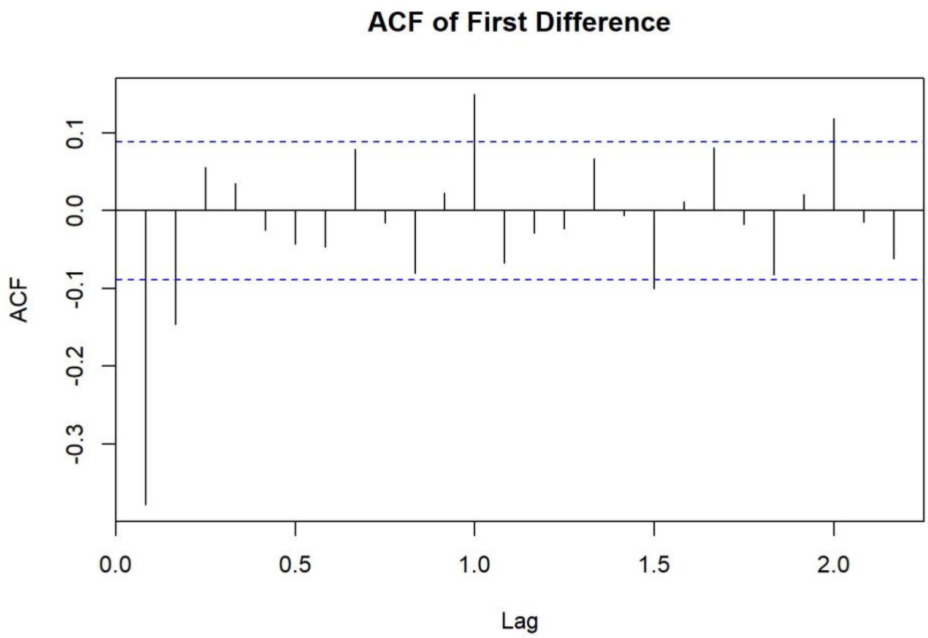
To address the non-stationarity issues in our data, we then apply the differencing technique to the time series and produce visual diagnostics (i.e., time series plot, ACF plot, and PACF plot). Specifically, we performed the first difference of our time series.

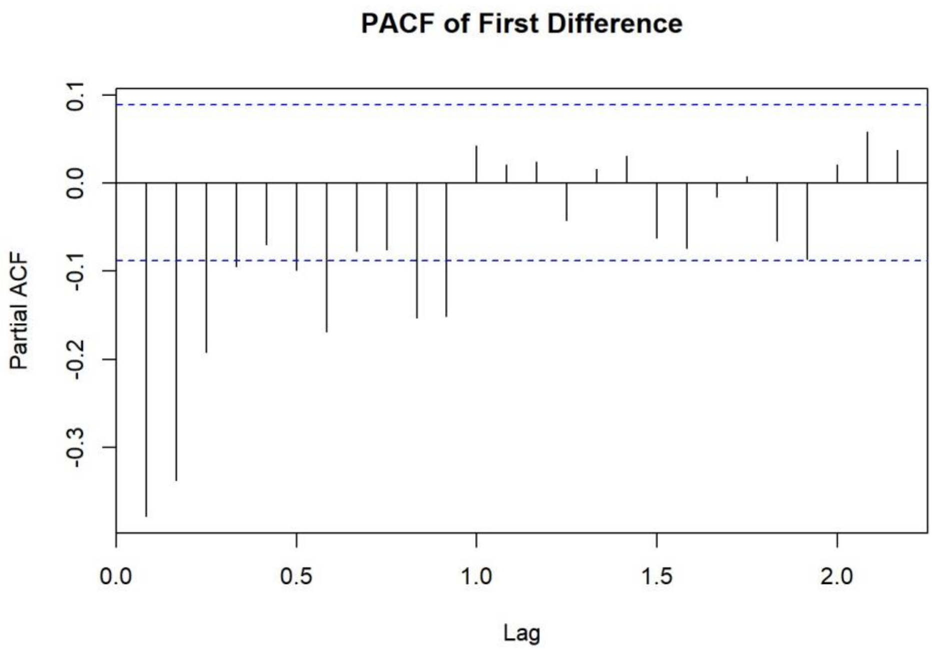
Time Series Plot - First Difference



Compared to the previous time series plot, this graph shows a constant mean and a relatively constant variance over time.

Plots of ACF and PACF - First Difference





After differencing, the ACF and PACF plots all show a quick decay as the lag becomes larger. From the plots, we still see periodic patterns of significant spikes at multiples of lag 12, suggesting the existence of remaining seasonality in our data. We suspect that our model should include an MA and an AR process, along with a difference term of d = 1. To ensure that our selection is reliable, we computed the EACF to assist us.

EACF Table

```
AR/MA
  0 1 2 3 4 5 6 7 8 9 10 11 12 13
0 x x o o o o o o o o o  x  o  o
1 x x o o o o o o o x o  x  o  o
2 x o x o o o x o o o o  x  o  o
3 x x o o o o o o o o o  x  o  o
4 x x o o o o o o o o o  x  o  o
5 x x o o x x o o o o o  x  o  o
6 x o x x o x x o o o o  x  o  o
7 x x x x o o o x o o o  o  o  o
```

Unfortunately, the EACF table presents multiple candidates for the "best" seasonal ARIMA model so we cannot take simple guesses of the non-seasonal component of the ARIMA orders or the seasonal part of the ARIMA orders from the analysis we did above. Therefore, we use the auto.arima() function from the forecast packages in R to select the model, which serves as a starting point for us to find the most appropriate model.

Seasonal ARIMA Model

```
Series: oat_dt_tsa
ARIMA(3,1,1)(1,0,2)[12]

Coefficients:
         ar1     ar2     ar3      ma1      sar1    sma1    sma2
      0.2775  0.0539  0.1486  -0.9769  -0.1123  0.3195  0.1724
s.e.  0.0474  0.0484  0.0477   0.0131   0.3348  0.3304  0.0724

sigma^2 = 2231119:  log likelihood = -4283.2
AIC=8582.4   AICc=8582.7   BIC=8615.97
```

Based on the output, the model is ARIMA(3, 1, 1)*ARIMA(1, 0, 2) s = 12. Notably, the z-statistics of the estimated sar1 coefficient is insignificant with a value of approximately 0.96. Hence, we consider the seasonal component with orders of P = 0, D = 0, and Q = 2.

Estimations:

We examine four additional seasonal ARIMA models that we consider to be a good fit for our time series data.

(1). Maximum Likelihood Estimates for a Simulated ARIMA(3, 1, 2)*ARIMA(1, 0, 2) Model:

```
Coefficients:
         ar1      ar2     ar3      ma1     ma2     sar1    sma1    sma2   drift
      0.5817  -0.0265  0.1278  -1.3005  0.3006  -0.0826  0.2935  0.1700  5.1137
s.e.  0.2550   0.0898  0.0618   0.2567  0.2566   0.3325  0.3301  0.0761  1.3396

sigma^2 = 2206519:  log likelihood = -4280.81
AIC=8581.62   AICc=8582.08   BIC=8623.58
```

In this model, by computing corresponding z-statistics, we observe insignificant estimated coefficients for ar2, ma1, ma2, sar1 and sma1.

(2). Maximum Likelihood Estimates for a Simulated ARIMA(3, 1, 1)*ARIMA(0, 0, 2) Model:

```
Coefficients:
         ar1     ar2     ar3      ma1    sma1    sma2   drift
      0.2895  0.0627  0.1620  -1.0000  0.2147  0.1582  5.1114
s.e.  0.0451  0.0469  0.0449   0.0077  0.0490  0.0492  1.2718

sigma^2 = 2202944:  log likelihood = -4281.48
AIC=8578.97   AICc=8579.26   BIC=8612.54
```

In this model, by computing corresponding z-statistics, we observe insignificant estimated coefficients for ar2.

(3). Maximum Likelihood Estimates for a Simulated ARIMA(3, 1, 1)*ARIMA(1, 0, 1) Model:

```
Coefficients:
         ar1     ar2     ar3      ma1    sar1     sma1   drift
      0.2894  0.0873  0.1808  -1.0000  0.9105  -0.7744  5.2033
s.e.  0.0449  0.0467  0.0447   0.0054  0.0384   0.0584  1.7458

sigma^2 = 2135794:  log likelihood = -4274.27
AIC=8564.53   AICc=8564.83   BIC=8598.1
```

In this model, by computing corresponding z-statistics, we observe insignificant estimated coefficients for ar2.

(4). Maximum Likelihood Estimates for a Simulated ARIMA(2, 1, 1)*ARIMA(1, 0, 2) Model:

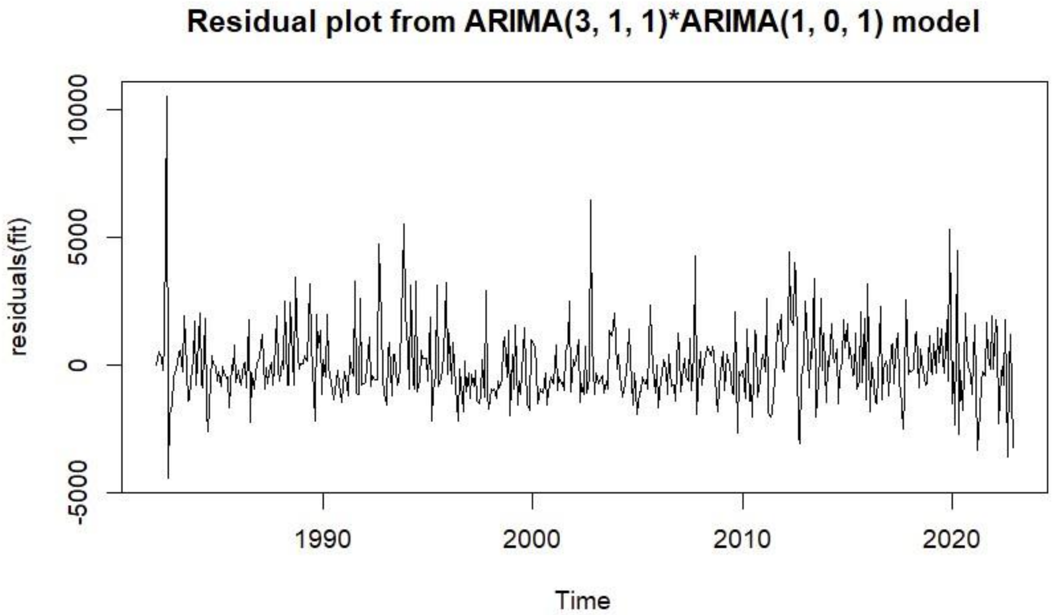
```
Coefficients:
         ar1     ar2      ma1     sar1    sma1    sma2   drift
      0.2801  0.0913  -0.9747  -0.3256  0.5333  0.2055  5.0944
s.e.  0.0563  0.0562   0.0318   0.4302  0.4159  0.0639  3.8983

sigma^2 = 2269065:  log likelihood = -4287.47
AIC=8590.95   AICc=8591.25   BIC=8624.52
```

In this model, by computing corresponding z-statistics, we observe insignificant estimated coefficients for ar2, sar1 and sma1.

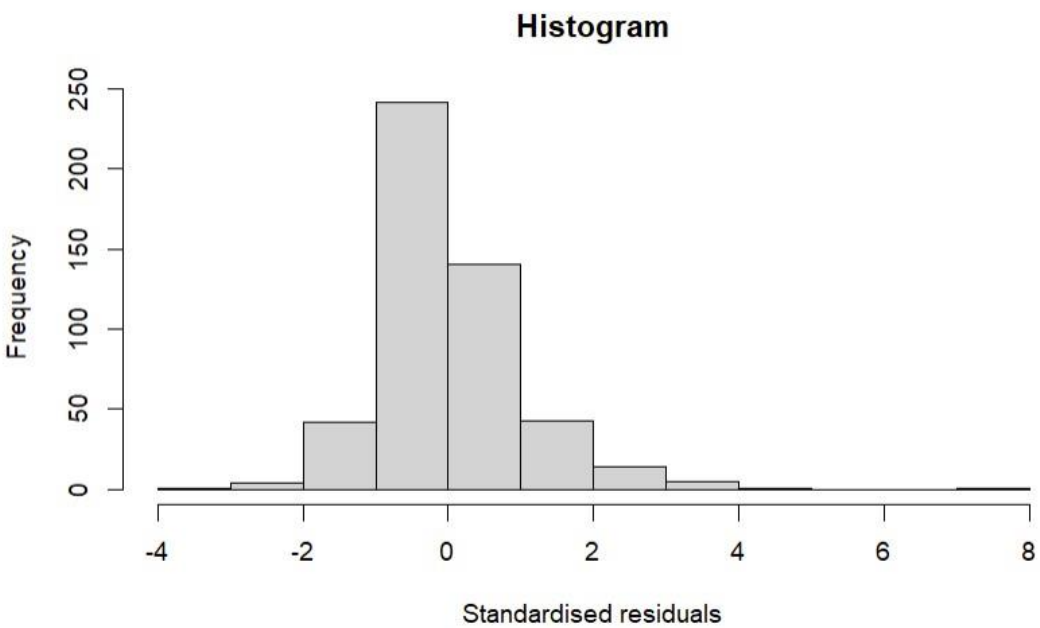
After comparing the results generated from each model, it is reasonable to conclude that the seasonal ARIMA (3,1,1) × (1,0,1) [12] model should be the best-fitted model since its AIC and BIC are the smallest ones.

Model Diagnostic

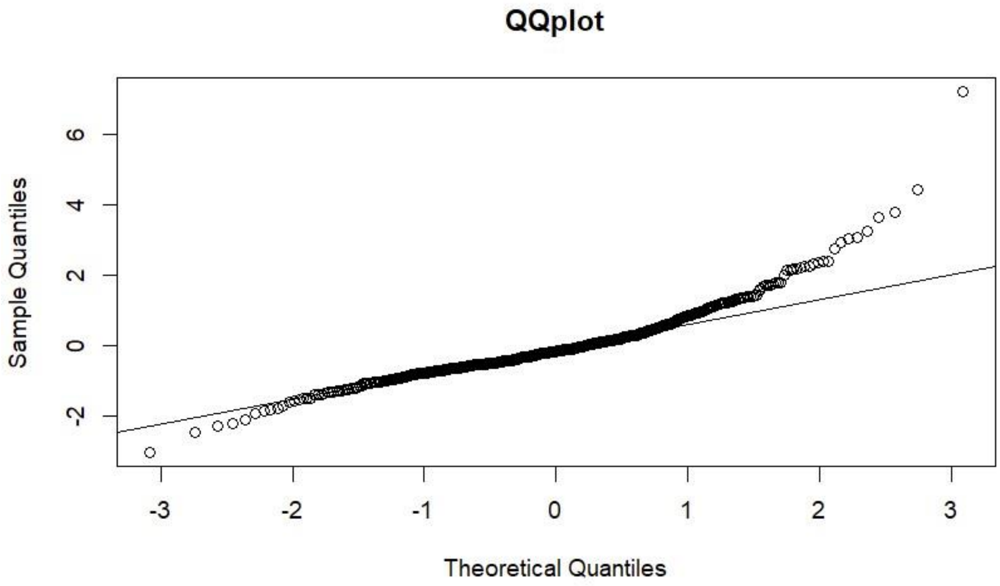
We use diagnostic analysis to test the goodness-of-fit of our selected seasonal ARIMA model. To illustrate, we plot residuals against time to check for potential violations of the zero mean and equal variance assumptions. In addition, we check if our residuals follow a normal distribution by using histograms and Q-Q plots. Finally, we plot a correlogram to examine the autocorrelation structures of our residuals.



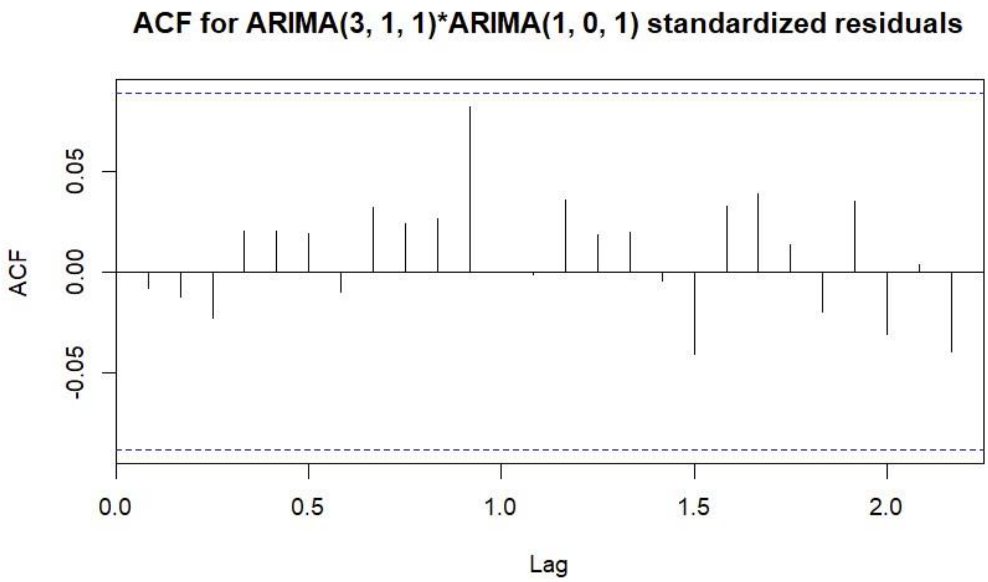
No discernible or organized pattern emerges over time from the residuals. It looks like there is no drift in the residuals as they remain close to zero. As spikes that contrast sharply with the overall amount of noise, there appear to be a few possible outliers or extreme values.



The histogram of the residuals exhibits right skewness with a long tail at the far end. This may suggest that the residuals might not be perfectly normally distributed.



The Q-Q plot of the residuals shows minor deviations from normality at the lower end. In contrast, the deviations at the right end are more significant. Although this may indicate the violation of the normality assumption of our model, we argue that the impact of slight non-normality on our inference could be negligible because other aspects of our diagnostic analysis are satisfactory. Moreover, the departures from normality are relatively limited to the extreme tails, leaving the central portion of the distribution unaffected. As a result, we conclude that the residuals of the seasonal ARIMA model follow a normal distribution (though not perfectly).



Our correlogram for the selected seasonal ARIMA model (i.e., ARIMA (3,1,1) × (1,0,1) [12]) does not exhibit significant autocorrelations at non-zero lags. In other words, the residuals of our model appear to be independent without systematic patterns, behaving like white noise.

Forecasting

The 90% confidence limit for the next 12 months of oats production in tonnes

	Point Forecast	Lower 90%	Higher 90%
Jan 2023	2946.635	533.4517	5359.819
Fed 2023	3928.509	1409.8108	6447.207
Mar 2023	3391.353	834.4150	5948.290
Apr 2023	3504.730	868.0404	6141.420
May 2023	3178.078	513.9022	5842.255
Jun 2023	3468.006	789.6783	6146.333
Jul 2023	3271.968	581.1891	5962.746
Aug 2023	3794.588	1096.4461	6492.731
Sep 2023	3563.520	860.5004	6266.539
Oct 2023	3874.305	1167.4506	6581.160
Nov 2023	4166.130	1456.3709	6875.888
Dec 2023	4237.098	1525.0193	6949.177

The table offers point forecasts and 90% confidence interval forecasts for the upcoming 12 months, derived from a fitted ARIMA (3, 1, 1)×(1,0,1) [12] model. Notably, the forecasts for this period unveil a nuanced pattern: an initial month with a gradual upward trend, succeeded by a downward trend in the subsequent four months, and ultimately, a resumption of the upward trend in the final months. This suggests an overall expectation of increased oat production over the next 12 months, despite the presence of a temporary downward trend within this period.

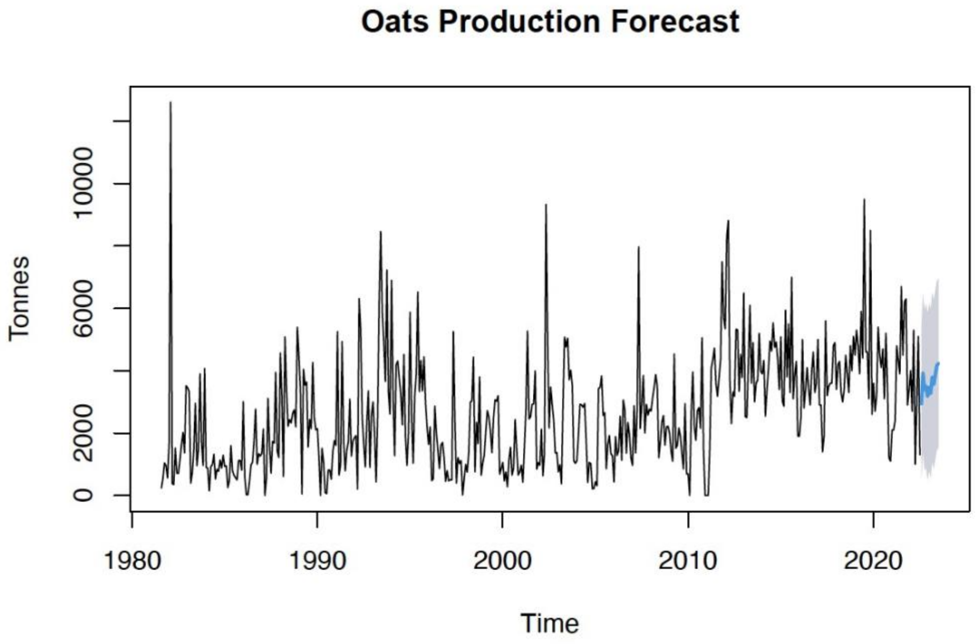
The 50% confidence limit for the next 12 months of oats production in tonnes

	Point Forecast	Lower 50%	Higher 50%
Jan 2023	2946.635	1957.084	3936.187

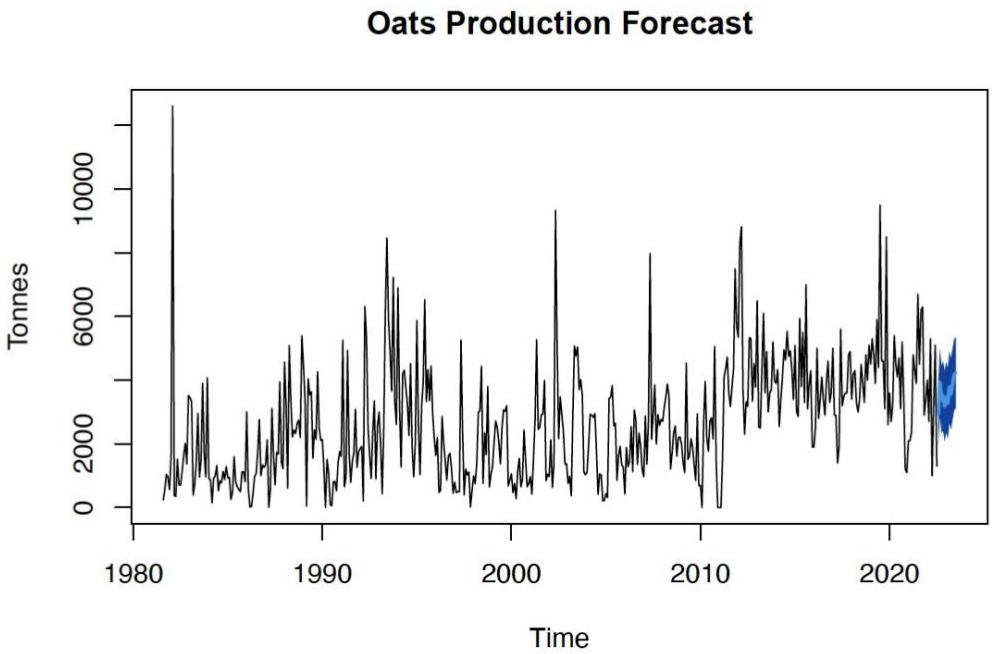
Fed 2023	3928.509	2895.690	4961.328
Mar 2023	3391.353	2342.853	4439.852
Apr 2023	3504.730	2423.527	4585.933
May 2023	3178.078	2085.605	4270.552
Jun 2023	3468.006	2369.729	4566.282
Jul 2023	3271.968	2168.585	4375.350
Aug 2023	3794.588	2688.186	4900.990
Sep 2023	3563.520	2455.118	4671.922
Oct 2023	3874.305	2764.331	4984.280
Nov 2023	4166.130	3054.964	5277.295
Dec 2023	4237.098	3124.981	5349.215

The table offers point forecasts and 50% confidence interval forecasts for the upcoming 12 months, derived from a fitted ARIMA (3, 1, 1)×(1,0,1) [12] model. The forecasts exhibit similar patterns as described by the predictions made with a 90% confidence limit. Still, this suggests that, for the next 12 months, the overall oat production in BC is expected to grow with some temporary declines.

Plot of forecasting of 90% confidence interval:



The plot of forecasting of 50% confidence interval:



These two charts contain observations from the past 40 years as well as forecasts for the coming year. Despite the differences in confidence levels, both charts agree that oat production is expected to increase next year, albeit with some fluctuations in the forecast trajectory.

Summary

This project focuses on conducting a comprehensive time series analysis of monthly oat producer delivery data in British Columbia, ranging from January 1982 to December 2022. The primary goal is to select and fit an accurate forecasting model, utilizing ARIMA methodology, to predict future oat production in the province.

Our data was sourced from Statistics Canada, covering monthly producer deliveries of major grains, with an emphasis on the production of oats. The dataset excludes observations outside the specified time frame (i.e., observations prior to 1982 and observations after 2022) to ensure accurate handling of potential seasonality in the analysis.

To ensure the proper application of an ARIMA model with seasonal components, our model fulfills the following key assumptions:

1. Weak stationarity - the time series exhibits constant mean and variance over time.
2. Independent observations – the autocorrelation values of the observations do not exhibit observable patterns that comprise our inferences.
3. Homoscedasticity – the variances of residuals of our fitted seasonal model remain constant over time.
4. Normality of residuals – the residuals follow approximately a normal distribution.
5. Large sample size – our fitted seasonal ARIMA model used a sufficient number of data points (monthly data collected for over four decades) to produce reliable parameter estimates.

The methods employed to select and fit a seasonal ARIMA model for forecasting future oats production in BC begin with visualization of the time series data, allowing for the identification of trends and seasonality. Then we proceeded with caution to determine the proper orders of the seasonal ARIMA model with the aid of ACF, PACF, and EACF plots after differencing the model to ensure that we meet the stationarity assumption. Subsequently, we conducted parameter estimation with the help of statistical software to optimize likelihood functions. We also conducted the diagnostic analysis to validate the chosen SARIMA model. Finally, we made predictions with two different prediction limits to examine the future changes in the quantity of oats produced in BC.

The findings of our analysis suggest a positive outlook for future oats production in 2023, reflected by an overall trend of growth. Our forecasts also acknowledge the possibility of temporary declines along the way. By recognizing the dynamic nature of the oats production forecast, the BC government can formulate policies to promote a resilient and adaptable environment for local farmers in the face of uncertainties.

Reference

Statistics Canada. (2023). Table 32-10-0351-01 International merchandise trade monthly price and volume. Retrieved Novermber 29, 2023, from https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=3210035101