A logo for college computing

Description automatically generated

**Assessment Cover Page**

|  |  |
| --- | --- |
| *Student Full Name* | Denisse Guadarrama Tinoco |
| *Student Number* | 2024154 |
| *Module Title* | Strategic Thinking |
| *Assessment Title* | CA 2 – Drought in Mexico |
| *Assessment Due Date* | 19th May 2024 |
| *Date of Submission* | 19th May 2024 |

**Declaration**

By submitting this assessment, I confirm that I have read the CCT policy on academic misconduct and understand the implications of submitting work that is not my own or does not appropriately reference material taken from a third party or other source.

I declare it to be my own work and that all material from third parties has been appropriately referenced.

I further confirm that this work has not previously been submitted for assessment by myself or someone else in CCT College Dublin or any other higher education institution.

Contents

[Introduction: 1](#_Toc167027504)

[Project schedule: 2](#_Toc167027505)

[Objectives: 3](#_Toc167027506)

[Problem Definition: 3](#_Toc167027507)

[Scope: 3](#_Toc167027508)

[Data Sources: 3](#_Toc167027509)

[Ethical Considerations: 4](#_Toc167027510)

[Data Preprocessing: 5](#_Toc167027511)

[Modelling: 9](#_Toc167027512)

[Results 26](#_Toc167027513)

[References 28](#_Toc167027514)

# Introduction:

“Mexico City is running out of water due to a drought crisis affecting 22 MILLION as 'Day Zero' when its supply won't be enough for residents is just weeks away” (Torres, 2024)

Currently at least 70% of Mexico suffers from water shortage, indicators show that the drought is increasing by 2024, approaching one of its worst droughts on record.

This event is not something new for the country, during 2020 and 2021 more than half of the Mexican territory was affected by the second most severe drought recorded since 2003, the first happened in 2011, according to the drought monitor of the National Water Commission (CONAGUA)

Droughts in Mexico are a problematic that is happening now and more and more frequently. It is a country with a lot of biodiversity and extensive hydrography; however, the volume of water decreases considerably between the dry season and the rainy season, due to the climate change and human actions.

We will be able to find a lot of information about water in Mexico, the availability in the country, rainfall, drought records by severity, among many other databases that are within our reach and that are free to use.

Will analysing this data and finding a model that helps project future droughts be able to convince citizens about the importance of water care? The answer is, in an ideal world it would be enough but the reality is different. Therefore, with the information obtained from this Project, it is expected to project the situation of Mexico in the future on droughts and water availability and thus be able to make decisions and actions based on this situation.

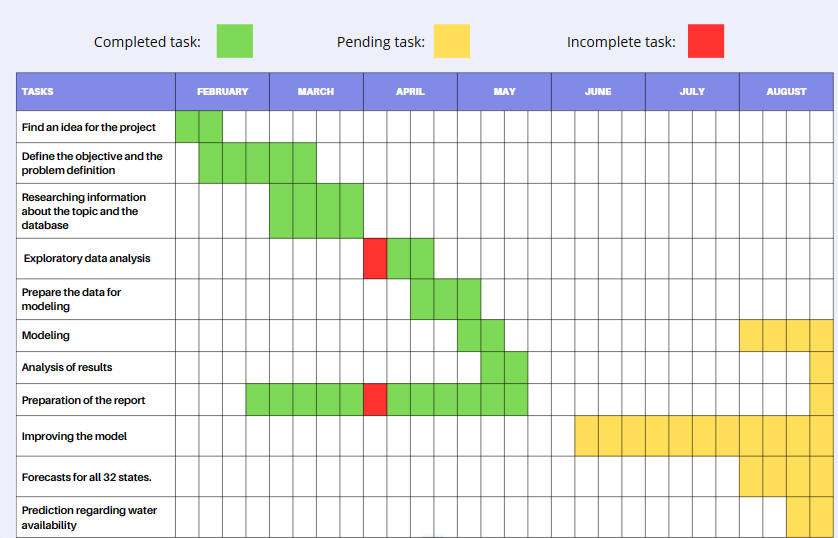
In this project you will be able to find the current drought situation in Mexico, how they are classified by facets and what are their characteristics. It will also be explained who are the organizations that record this information and the databases used.

Some factors identified by the UNAM (National Autonomous University of Mexico) that favour water scarcity will also be discussed, to find out if there is a relationship and to be able to predict the amount of water available by 2025 per capita.

The priority of this Project is to find the municipalities with the most severe droughts by 2025 and the month in which this happens, using machine learning models that help predict them. It will be explained what models were used and how the databases were used.

Water is a valuable and important resource and making use of the knowledge learned to help visualize the situation in Mexico is a first step to making change.

# Project schedule:



# Objectives:

The main objective of this Project is to present information on the situation of droughts, identifying through predictions which states in Mexico will suffer from the most severe droughts and the period of time in which this happens in 2025, as well as classifying the level of drought by each state

As a preliminary approach, one state from each region will be compared to observe the differences in drought levels across the country and to see how drought levels can vary depending on the region.

Additional predictions will be presented about the availability of water by region or by State in 2025, taking into account factors that could be relevant to the cause of water shortages.

# Problem Definition:

Due to the Environmental problem that Mexico suffers, severe droughts have occurred during the last 20 years with or without many actions to reduce it, this leaves us uncertain as to what the droughts will be like in 2025, therefore, to have a better visualization of the problem that is approaching, the levels of drought will be forecast by states, so the pertinent actions can be taken to solve or improve the situation, as well as inform the citizens since “Mexico must address the water problem with data and evidence” (Instituto Mexicano para la Competitividad A.C., 2023)

# Scope:

Water scarcity is a global problem, but for this Project we will only focus on the information obtained from Mexico.

The purpose of this Project is not to find the Source, diagnose the cause of droughts, or propose actions to solve it.

It is a Project that will provide information on what 2025 could be like in Mexico related to droughts. The following points list the outcomes that are expected to be achieved:

* Projection of the main states with severe drought D04 and D03 in Mexico (only with recorded historical data) biweekly
* Biweekly forecast by state of drought level in 2024 and 2025 (only with registered historical data)
* Including 3 representative databases that affect the available volume of water, such as population growth, historical rainfall, presence of agriculture in the state and water treatment, the relation between these factors and the drought will be sought and, if it exists, it will be used to project the amount of water available per state.

# Data Sources:

The sources where the databases will be obtained are two mainly:

* CONAGUA: It is an institution that is responsible for the preservation and administration of national waters for their sustainable administration.

Its databases are published on the Mexican government's website called Open Data and are free to use: [Datos Abiertos de México - datos.gob.mx](https://datos.gob.mx/libreusomx)

* INEGI: ‘It is an autonomous public body responsible for regulating and coordinating the National System of Statistical and Geographical Information, as well as for collecting and disseminating information about Mexico in terms of territory, resources, population and economy’. (Instituto Nacional de Estadística y Geografía (INEGI), 1983)

The databases used are the following:

* History of the drought in each municipality

Source: CONAGUA

This database is our main source of information, they have daily and biweekly records since 2003.

* Population growth

Source: INEGI

* Rainfall

Source: CONAGUA

Behaviour of monthly average rainfall at the state and national level from 1985 to date, measured through conventional and automatic stations

* Agricultural sector

Source: INEGI

Number of agricultural activities in the country by state.

* Water availability

Source: CONAGUA

Average water availability per capita.

# Ethical Considerations:

This project is carried out during elections in Mexico; therefore, the use of the data and its interpretation must be neutral, this means not representing an official Government position.

Because it is a current problem, the data must be used with discretion, not used to deceive or confuse the population by changing the original meaning of the information and its veracity.

The information presented must be impartial, honest and transparent in order to optimally aid in decision-making.

On the other hand, there is a bias in data collection because the data is collected every 15 days, but it is not known for certain whether it is collected at the same time.

However, this information is collected through the interpretation of various drought indices and indicators, and through a consensus that determines the regions affected by drought. Since this information is used and collected by the government of Mexico, it can be assumed that the bias in the data collection and analysis is minimal.

# Data Preprocessing:

The data is in an xlsx file

The file contains information on drought levels, organized by city and state, reported since January 2003 and recorded biweekly.

file\_name = "sequia2023.xlsx"

5 rows × 379 columns

**Removing columns**

We will remove the unnecessary columns, as the majority of them are identification keys.

We will also remove the cities from each state, as our focus will be uniquely on states. We only need to know the state and the historical data of the droughts.

*df = df.drop(columns=["CVE\_CONCATENADA","CVE\_ENT","CVE\_MUN","NOMBRE\_MUN", "ORG\_CUENCA", "CLV\_OC","CVE\_CONC","CON\_CUENCA","CVE\_CONC"]*



Replacing the drought levels with numerical values as follows:

1. Minimum (10% water reduction) D0 = 1

2. Moderate (15%) D1= 2

3. Moderate (25%) D2= 3

4. Severe (40%) D3= 4

5. Critical (>40%) D4= 5

**Missing values:**

All NA values do not indicate missing data; rather, they signify that no drought was reported on those days. Therefore, we will replace them with 0.

*df\_replace=df\_replace.fillna(0*)

The 'Entidad' column contains the states of the country, repeated because the database is divided into cities. However, our objective is to determine the drought level by state. Therefore, the maximum value will be taken, meaning the highest drought level among all the cities belonging to a specific state will be assigned to the state. This approach is preferred because if there is a possibility that one city within the state has a high drought level, it is more prudent to consider the most critical situation rather than the average, for example.

*df\_replace["ENTIDAD"].unique()*

*Out[10]:*

*array(['Aguascalientes', 'Baja California', 'Baja California Sur',*

*'Campeche', 'Coahuila de Zaragoza', 'Colima', 'Chiapas',*

*'Chihuahua', 'Ciudad de México', 'Durango', 'Guanajuato',*

*'Guerrero', 'Hidalgo', 'Jalisco', 'Estado de México',*

*'Michoacán de Ocampo', 'Morelos', 'Nayarit', 'Nuevo León',*

*'Oaxaca', 'Puebla', 'Querétaro de Arteaga', 'Quintana Roo',*

*'San Luis Potosí', 'Sinaloa', 'Sonora', 'Tabasco', 'Tamaulipas',*

*'Tlaxcala', 'Veracruz de Ignacio de la Llave', 'Yucatán',*

*'Zacatecas'], dtype=object)*

*max = df\_replace.groupby('ENTIDAD').max()*

*new\_df* = *pd*.*DataFrame*(max)

We will create a new dataframe containing the maximum values and transpose it, so that each state is represented in a separate column.

*df\_trans = new\_df.T*

The date will be set as the index of our dataframe to enhance the visualization of our graphs and facilitate our predictive analyses.

*df\_new2=df\_new.set\_index(["index"])*

We have finished preparing our data and ensured that all columns contain values and are of the int32 type.

<class 'pandas.core.frame.DataFrame'>

DatetimeIndex: 370 entries, 2003-01-31 to 2023-12-31

Data columns (total 32 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 Aguascalientes 370 non-null int32

1 Baja California 370 non-null int32

2 Baja California Sur 370 non-null int32

3 Campeche 370 non-null int32

4 Chiapas 370 non-null int32

5 Chihuahua 370 non-null int32

6 Ciudad de México 370 non-null int32

7 Coahuila de Zaragoza 370 non-null int32

8 Colima 370 non-null int32

9 Durango 370 non-null int32

10 Estado de México 370 non-null int32

11 Guanajuato 370 non-null int32

12 Guerrero 370 non-null int32

13 Hidalgo 370 non-null int32

14 Jalisco 370 non-null int32

15 Michoacán de Ocampo 370 non-null int32

16 Morelos 370 non-null int32

17 Nayarit 370 non-null int32

18 Nuevo León 370 non-null int32

19 Oaxaca 370 non-null int32

20 Puebla 370 non-null int32

21 Querétaro de Arteaga 370 non-null int32

22 Quintana Roo 370 non-null int32

23 San Luis Potosí 370 non-null int32

24 Sinaloa 370 non-null int32

25 Sonora 370 non-null int32

26 Tabasco 370 non-null int32

27 Tamaulipas 370 non-null int32

28 Tlaxcala 370 non-null int32

29 Veracruz de Ignacio de la Llave 370 non-null int32

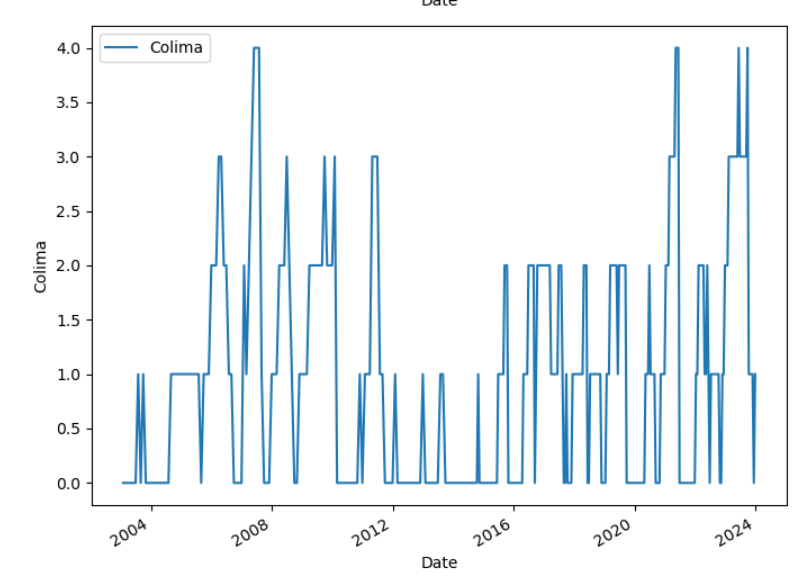
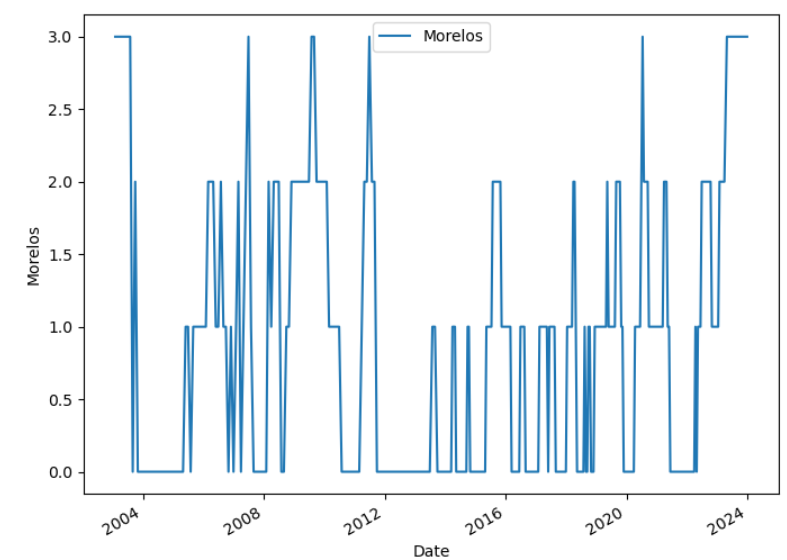
30 Yucatán 370 non-null int32

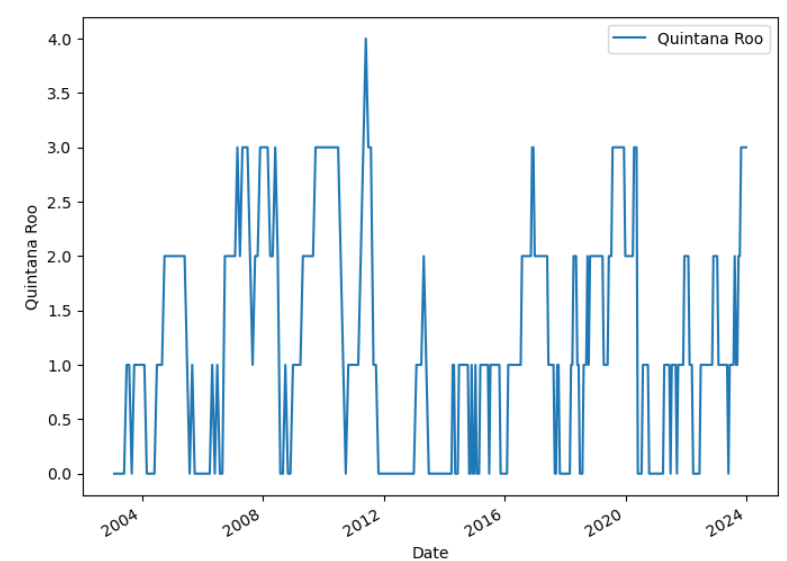
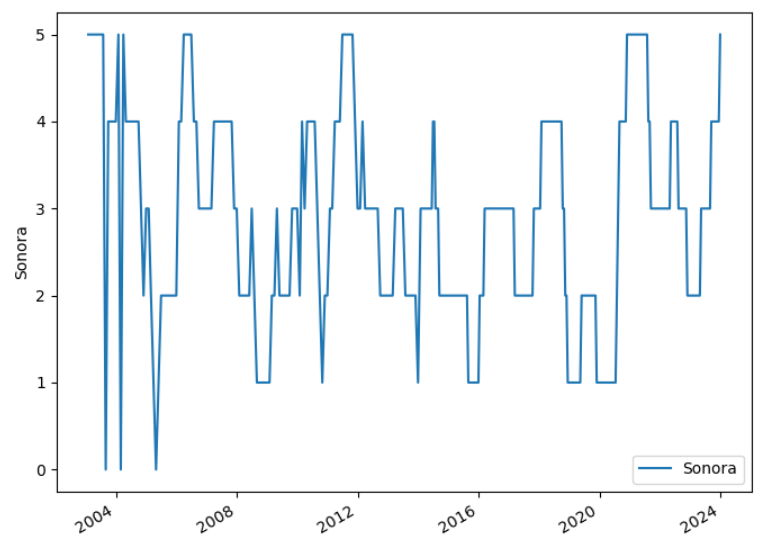
31 Zacatecas 370 non-null int32

dtypes: int32(32)

memory usage: 49.1 KB

Once our data is prepared, we will plot the drought levels for all states to observe their behavior since 2003, we can observe the 32 graphs in Python, only 4 will be displayed in this report

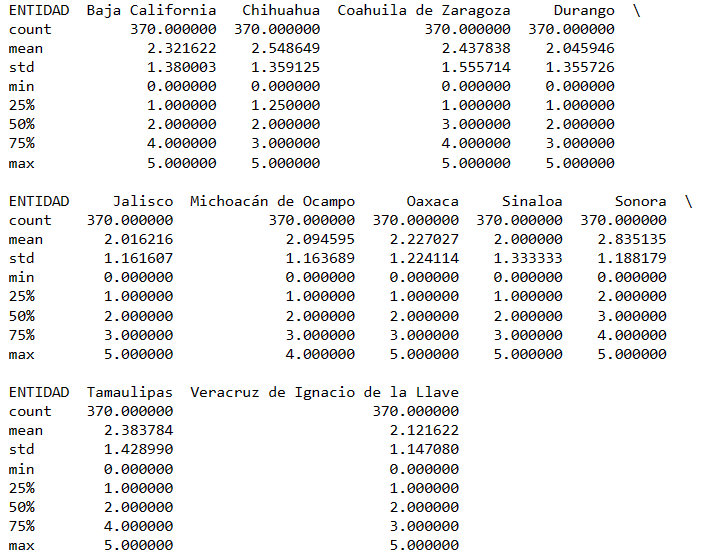
From these 4 graphs, we can observe that Sonora is the only one that has not decreased from level 1 since before 2008, and by the end of 2023, it reached level 5, which is the most critical level, similar to Morelos and Colima.



We observe that nearly all states exhibit maximum values of 5.

The country's average is 0.54, indicating that there are states with minimal drought occurrences. Conversely, states such as Sinaloa and Sonora exhibit the highest levels, with an average of 2.

The countries with averages above 2 are:



# Modelling:

We have our data prepared to apply the model for forecasting. For this project, we will utilize ARIMA, as it is a statistical forecasting model.

"This model comprises three components. The autoregressive element (AR) relates the current value to past (lagged) values. The moving average element (MA) assumes that the regression error is a linear combination of past forecast errors. Finally, the integrated component (I) indicates that the data values have been replaced with the difference between their values and the previous ones" (Amat Rodrigo & Escobar Ortiz, 2023)

In the model, the parameters p, d y q are utilized, each representing a component:

• p for the autoregressive element

• d number of differencing that have been performed

• q for the moving average element

It is not necessary to use all three components, we can have combinations where one or two components are 0, meaning they are not utilized.

The next step in the process will be to determine these parameters for the series.

**Stationarity (I)**

We need our data to be stationary to ensure that the forecasts are more reliable.

Therefore, we will conduct two tests for each column: the KPSS Test and the ADF Test. Both tests will compare the p-value against the alpha level (0.05) to determine whether the data series is stationary or not.

If:

p-value < α (0.05)

Kpss test = the series is non-stationary and thus needs to be differenced.

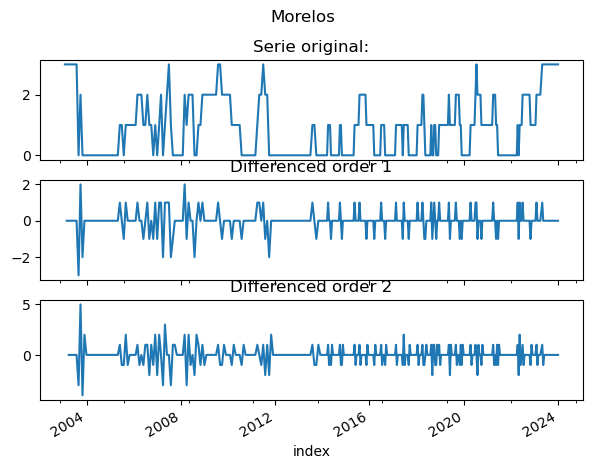
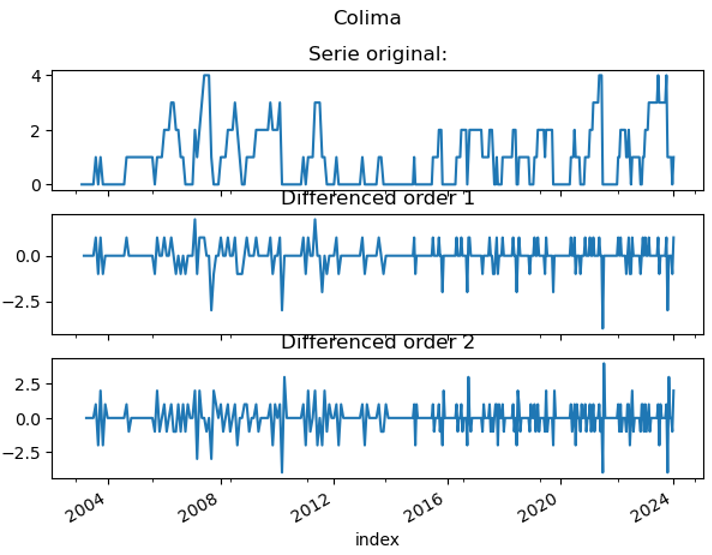
ADF test = the series is stationary.

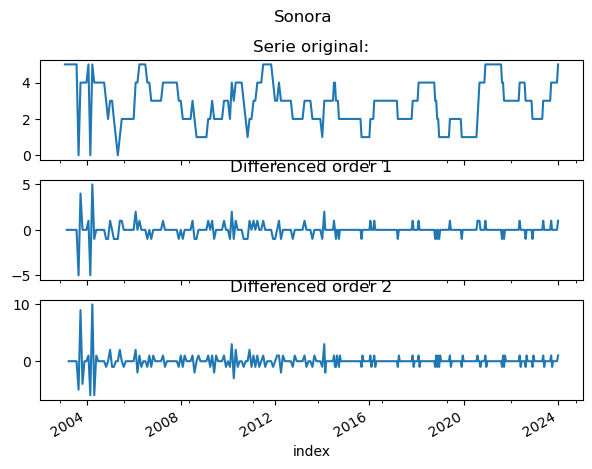
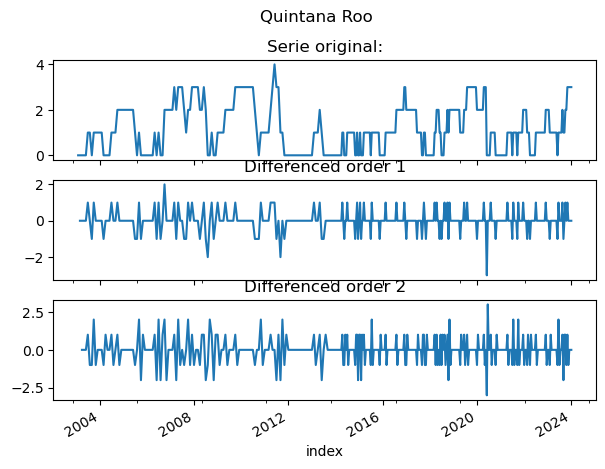
If the ADF indicates stationarity and the KPSS indicates non-stationarity, we might be in a situation of "difference stationarity." This means we need to apply differencing to make the series stationary.

The purpose of differencing is to make the time series stationary.

We need to difference the series until it becomes stationary, but always seeking the minimum differencing; otherwise, the series could be over-differenced which it will affect the model parameters.

We plotted the original and differentiated data once and twice for all 32 data series, but we will display the same 4.





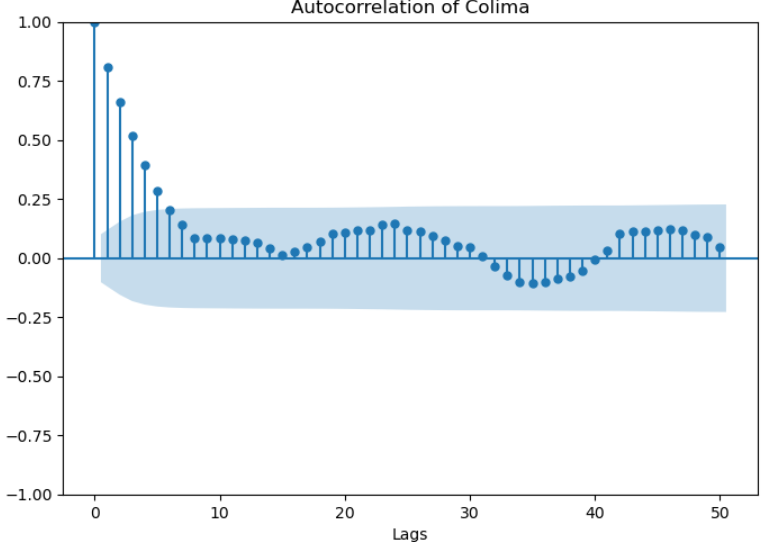
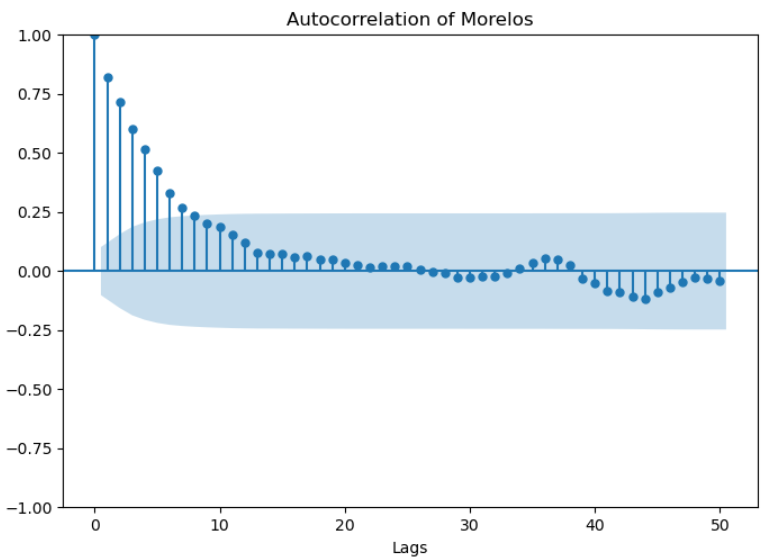
If the series is stationary, it is not differenced and d=0

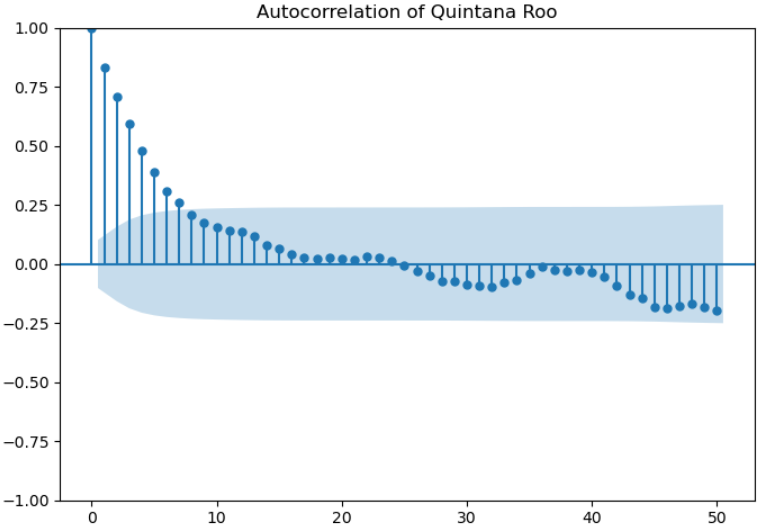
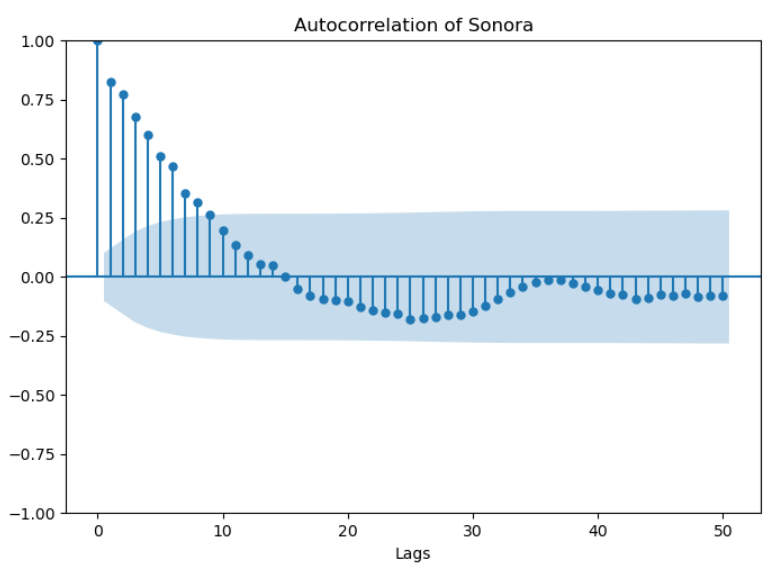
If the series is non-stationary, it is differenced twice, but we will take the first differencing if it passes both tests for stationarity, making d=1. If it needs another differencing, d=2.

The d values for each feature will be stored in value\_d, we will use this value for the model we are going to employ.

**Autocorrelation Function (ACF) for MA**

After having all our series stationary, we will plot the ACF to identify the value of q.

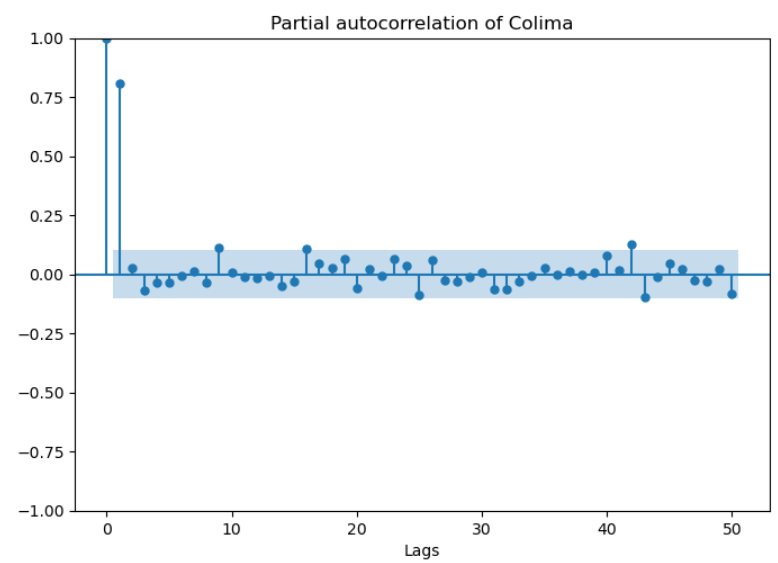
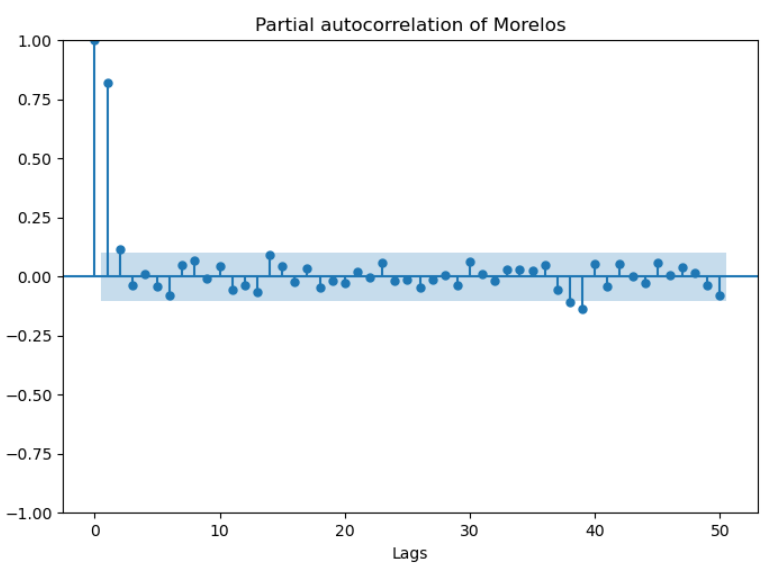
 

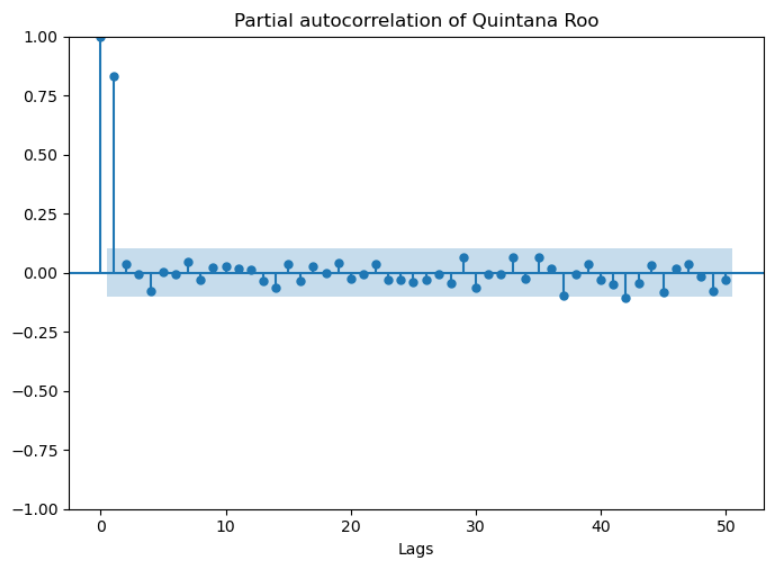
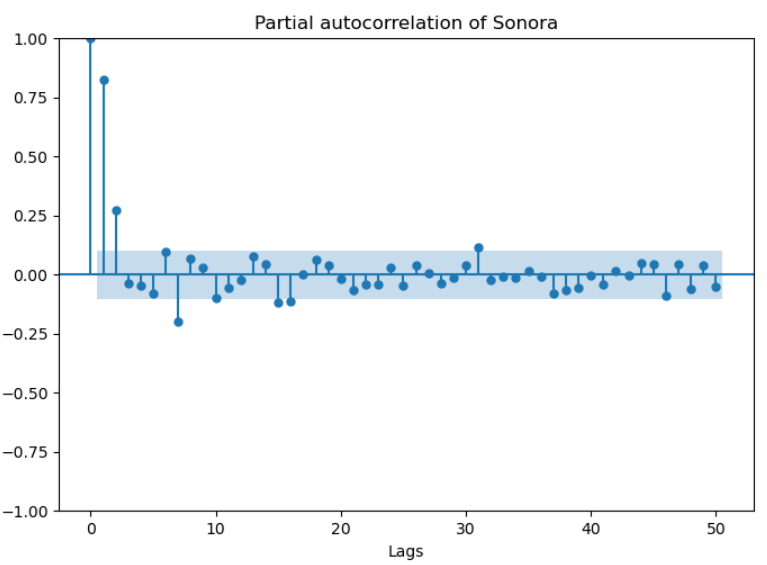
The ACF calculates the correlation between a time series and its lagged values. The lag at which the ACF falls provides an estimate of the value of q.

We will assign q for each series conservatively.

**Partial Autocorrelation Function (PACF) for AR**[**¶**](#Partial-Autocorrelation-Function-(PACF))

The PACF helps identify the value of p.

The PACF measures the correlation between a lagged value and the current value of the time series, taking into account the effect of intermediate lags. The lag at which the PACF cuts off provides an indication of the value of p.

We created a new dataframe to store the model parameters for each feature.

*valuesmodel.columns=names*

*new\_index=['d', 'q', 'p']*

*valuesmodel = valuesmodel.set\_index(pd.Index(new\_index))*

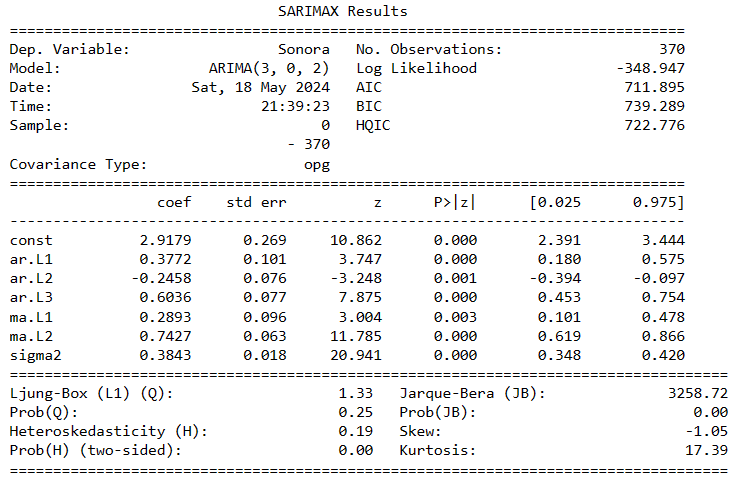
The next step is to apply the model. For this, we will randomly select 3 states, one from each region of Mexico, which are divided as follows:

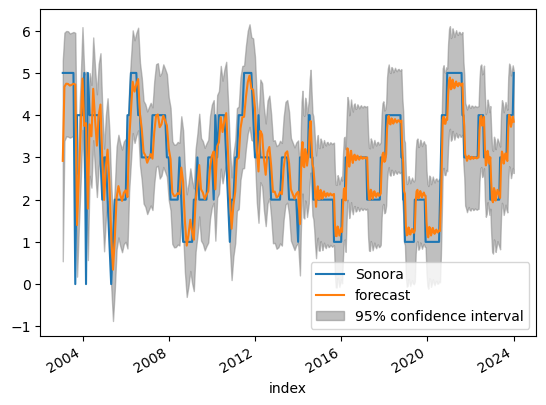
* Northern Region: Baja California, Baja California Sur, Sonora, Chihuahua, Coahuila, Nuevo León, Tamaulipas
* Central Region: Durango, Zacatecas, San Luis Potosí, Aguascalientes, Nayarit, Guanajuato, Querétaro, Michoacán, Estado de México, Hidalgo, Mexico City, Morelos, Tlaxcala, Puebla
* Pacific Region: Sinaloa, Colima, Jalisco, Guerrero
* Southern Region: Oaxaca, Chiapas, Tabasco, Campeche, Yucatan, Quinta Roo, Veracruz

These regions represent Mexico's geographical, climatic, economic, and cultural diversity, and each has its own unique characteristics and attractions.

**SONORA (Northern Region)**

We will utilize ARIMA with the values obtained from the analyses above.

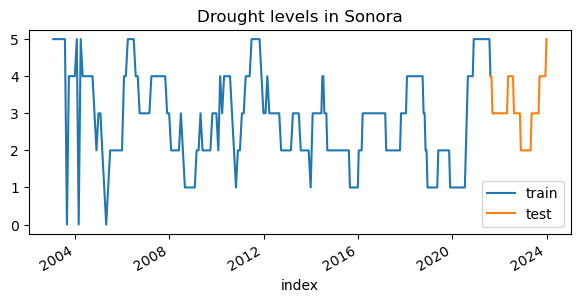


We plotted the model results against the actual results.

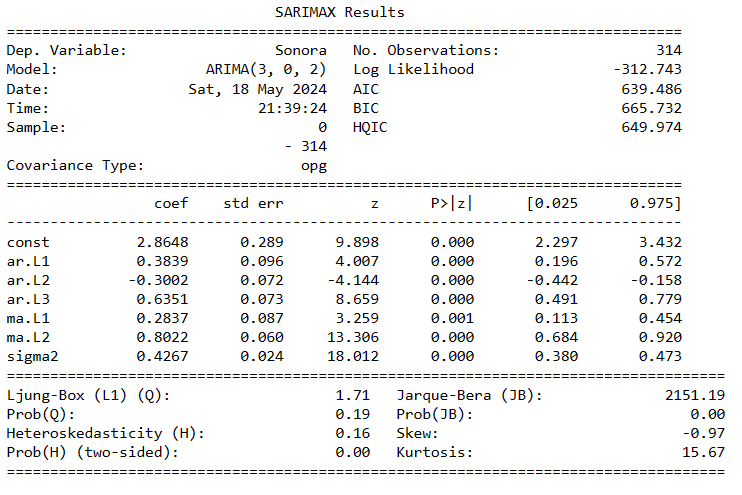
**Out-of-Time Cross validation**[**¶**](#Out-of-Time-Cross-validation)

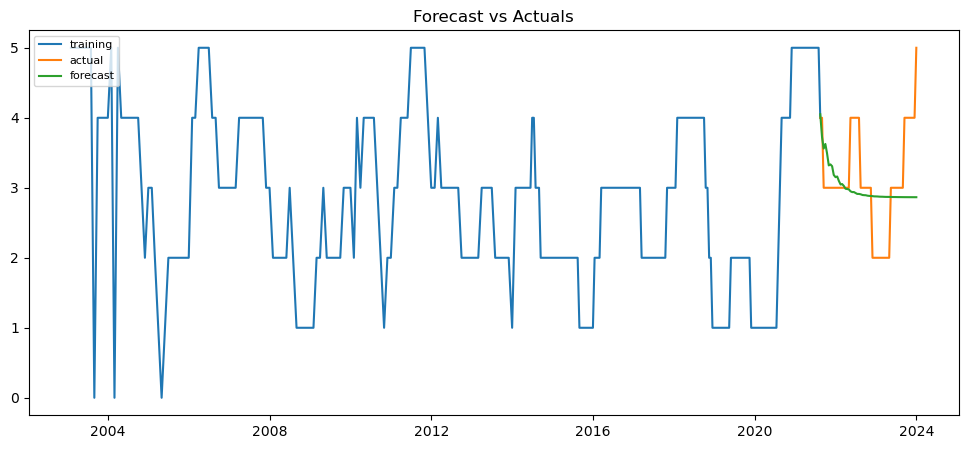
To do out-of-time cross-validation, we will create the training and testing dataset by splitting the time series into 2 contiguous segments, 85% and 15% respectively.

We will compare our forecast against the actual values.



We will apply the ARIMA model again using only the training data:





**Metrics:**[**¶**](#Metrics:)

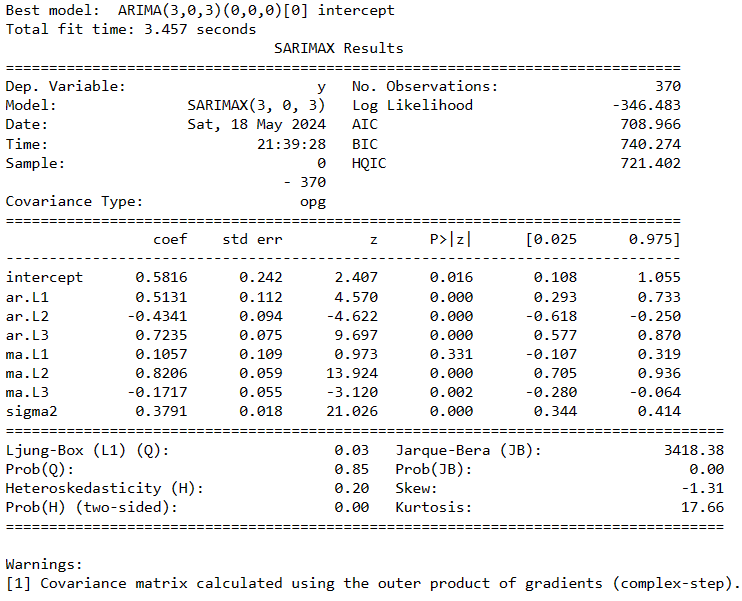
I will use the following metrics to assess the accuracy of the forecasts:

* 'mape': 0.1791524415519957,
* 'me': -0.11386948487869052,
* 'mae': 0.5402134212418244,
* 'mpe': 0.013912914648951577,
* 'rmse': 0.7222543523808019,
* 'corr': 0.13451354269182872,
* 'minmax': 0.1535632065257534

An approximate MAPE of 17.91% suggests the model is approximately 82.1% accurate in predicting the next 57 observations. However, upon inspection of the plot, the forecast does not closely resemble the actual data

**Auto Arima**[**¶**](#Auto-Arima)

We will use stepwise approach to search multiple combinations of p,d,q parameters and chooses the best model that has the least AIC.



The model summary shows a lot of information about the fitting process, we will use two metrics to compare the results of the two models with the parameter changes we made:

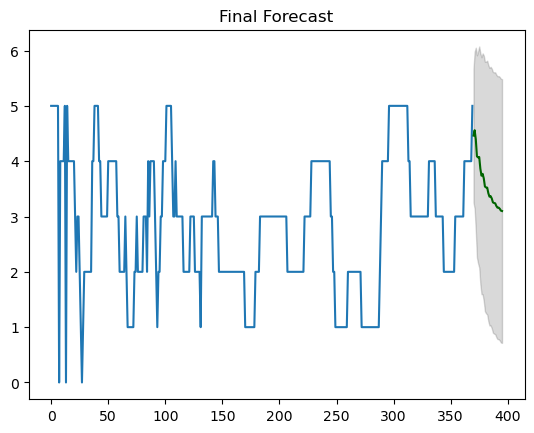
Log-Likelihood: It is a measure that explains how well the model fits the data. If it is negative, it indicates poor fit, and if it is close to 0, it indicates a better fit.

AIC: It is a goodness-of-fit metric. The lower it is, the better the model.

The AIC has reduced to 708 from 711 and the Log-likelihood improved from -348.947 to -346.483, however it is still a poor performance as the score remains negative.

**Forecast**:

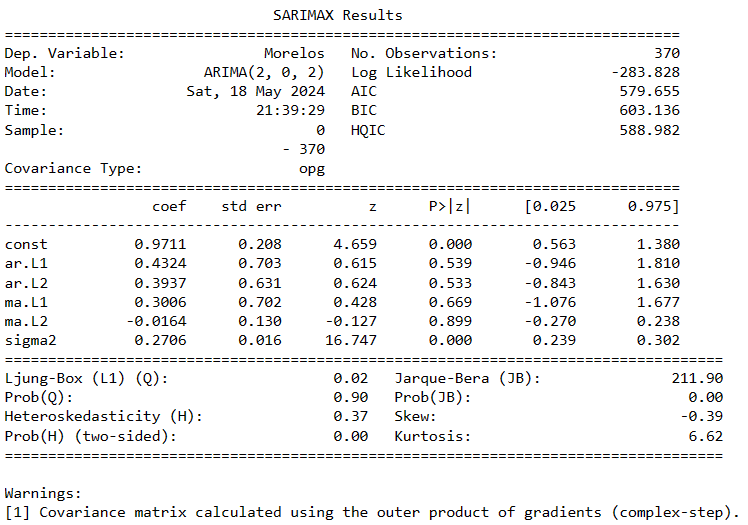
Using the latest model to predict the drought levels in Sonora for the 26 bi-weekly periods to cover all of 2024.

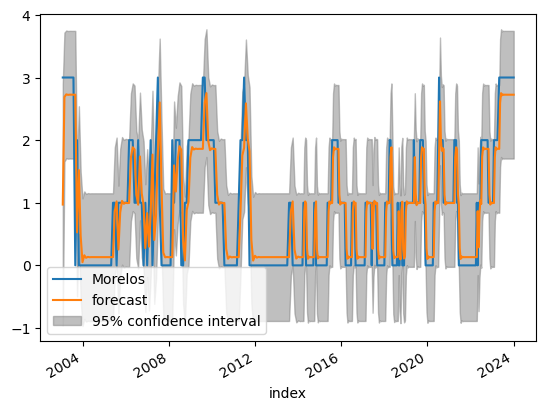


These are the results for 24 bi-weekly periods. We observe that we did not obtain good forecasts, as the values start high and then decrease. In contrast, the current drought levels fluctuate across all 4 levels throughout the year.

**MORELOS (Central Region)**:

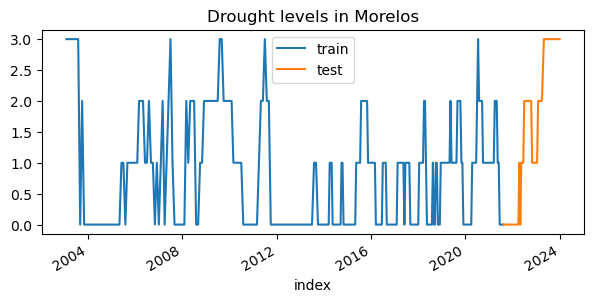
We will follow the same steps as above for this new dataset.

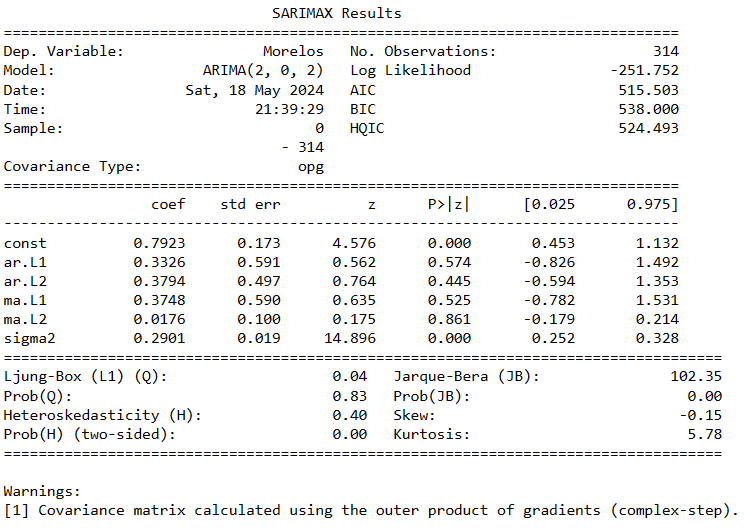


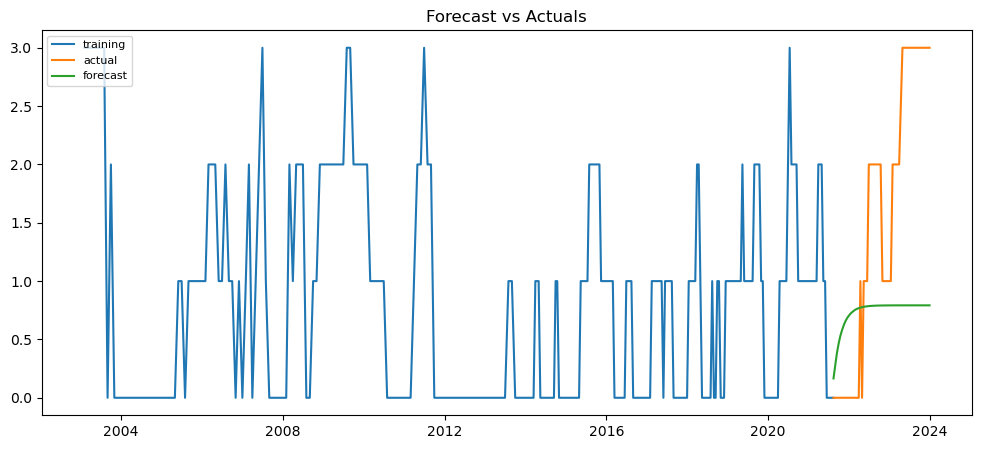
[**¶**](#MORELOS-(Central-Region))

**Out-of-Time Cross validation**[**¶**](#Out-of-Time-Cross-validation)

Splitting the data into train and test to do Out-of-Time Cross validation



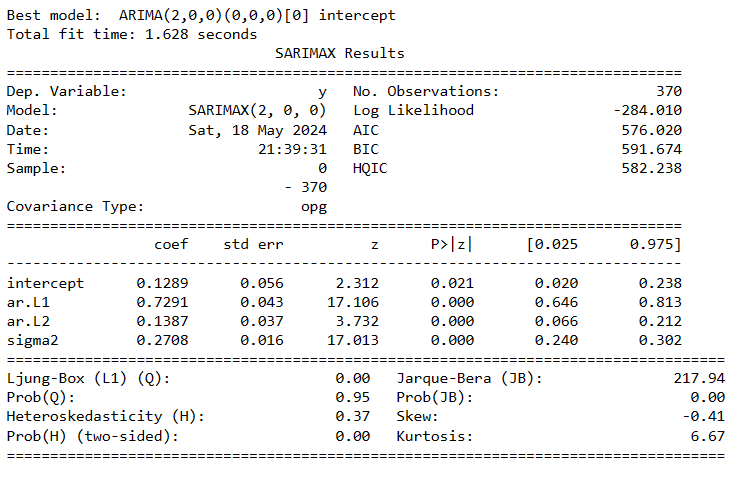




**Metrics**:

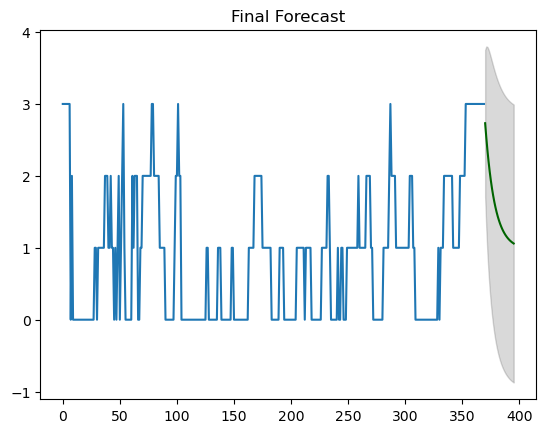
* 'mape': inf,
* 'me': -0.7906651679707022,
* 'mae': 1.1528780299039263,
* 'mpe': inf,
* 'rmse': 1.3838447749887062,
* 'corr': 0.546293951488008,
* 'minmax': 0.6930871372675147

**Auto Arima:**

****

The AIC has reduced to 576 from 579 and The Log-likelihood did not improve, the model did not fit the data correctly.

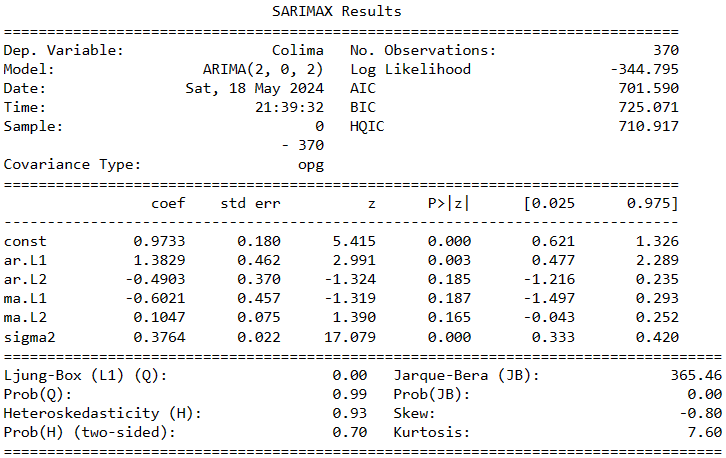
**Forecast:**

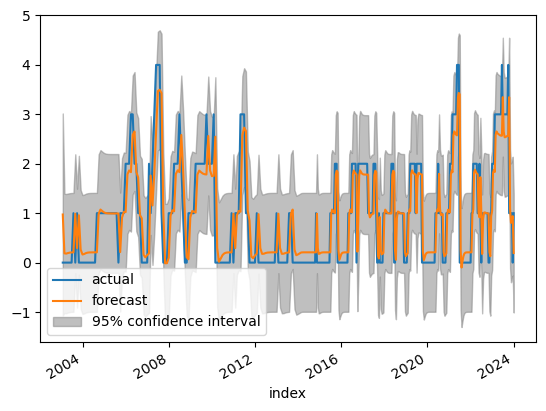


**COLIMA (Pacific Region)**

We will utilize ARIMA with the values obtained from the analyses above.

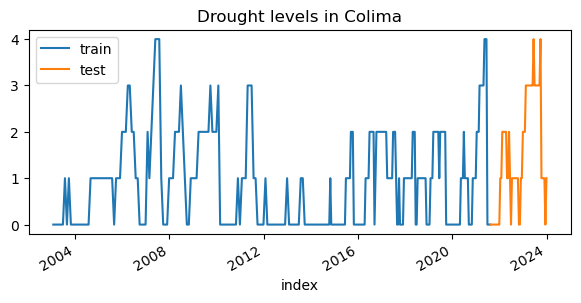
[**¶**](#COLIMA-(Pacific-Region))

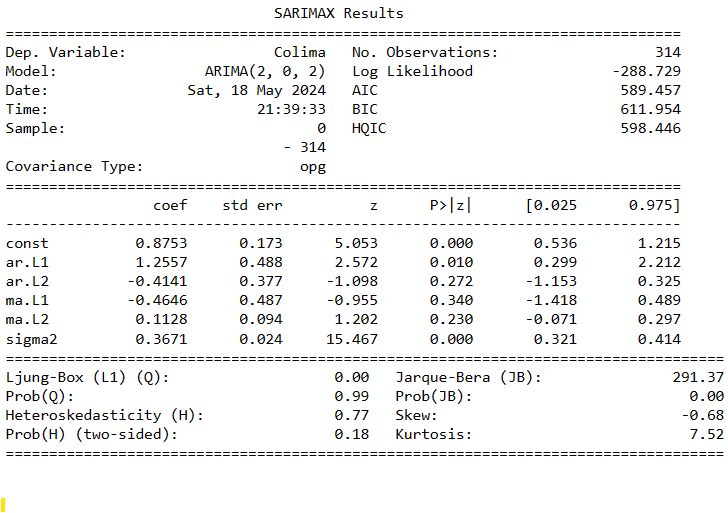


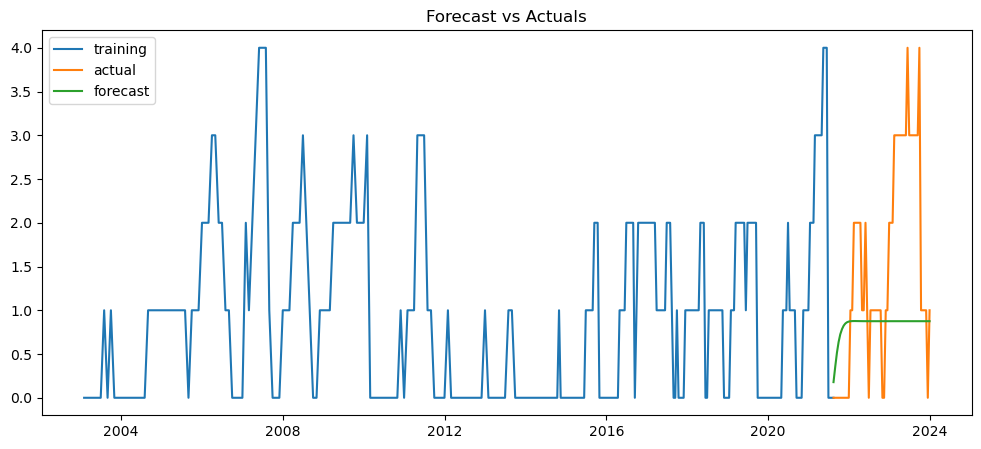


**Out-of-Time Cross validation**[**¶**](#Out-of-Time-Cross-validation)

Splitting the data into train and test to do Out-of-Time Cross validation



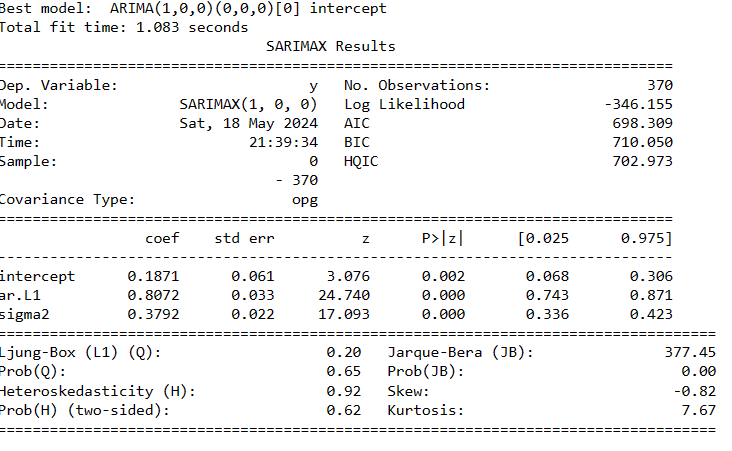




**Metrics**:

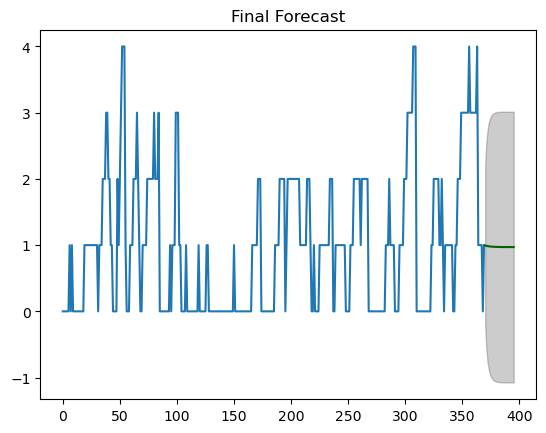
* 'mape': inf,
* 'me': -0.6358780861048975,
* 'mae': 0.9905869575407585,
* 'mpe': inf,
* 'rmse': 1.3100374871241802,
* 'corr': 0.3739537712024041,
* 'minmax': 0.5648269669296109

**Auto Arima**[**¶**](#Auto-Arima)



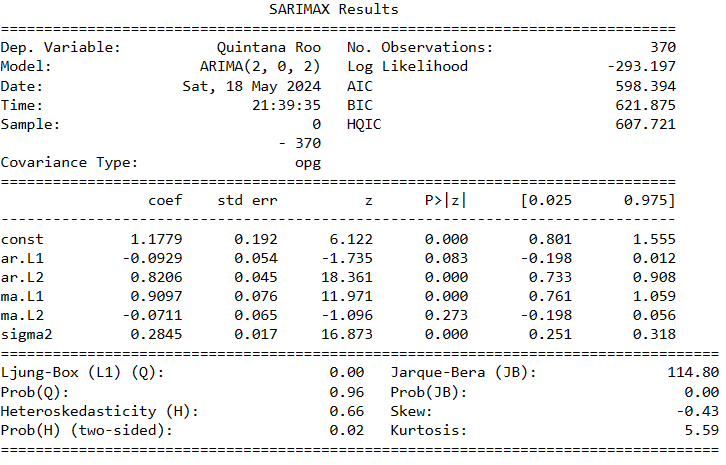
The AIC has reduced to 698 from 701.

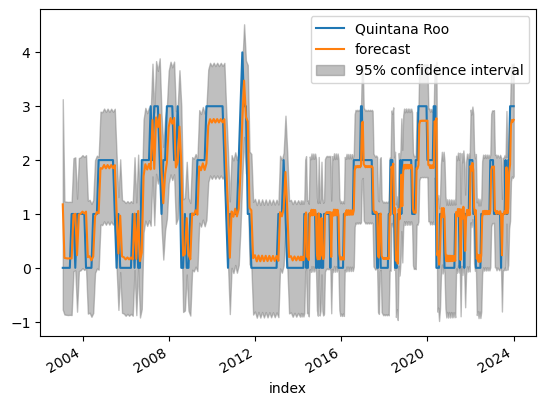
**Forecast**:



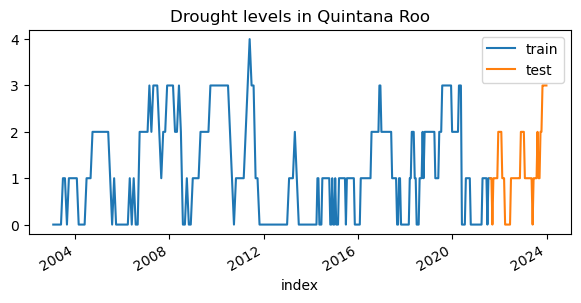
**QUINTANA ROO (Southern Region)**

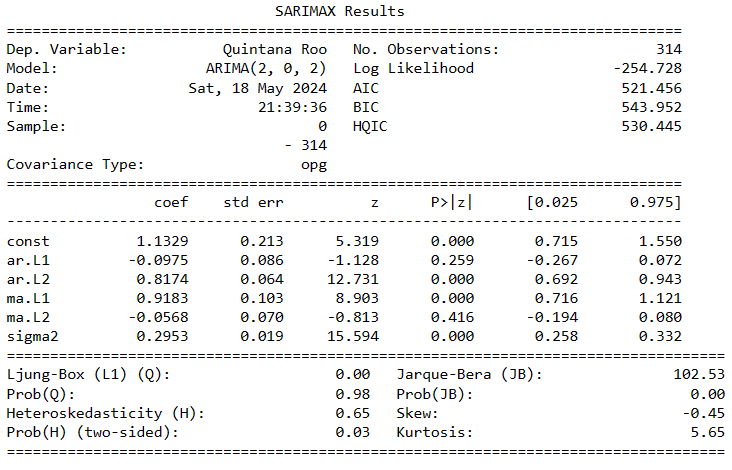
We will utilize ARIMA with the values obtained from the analyses above.

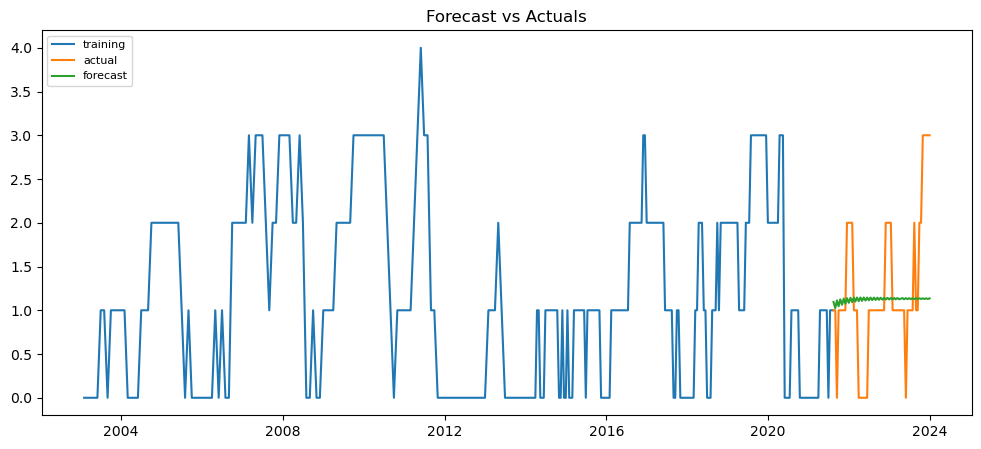




**Out-of-Time Cross validation**[**¶**](#Out-of-Time-Cross-validation)



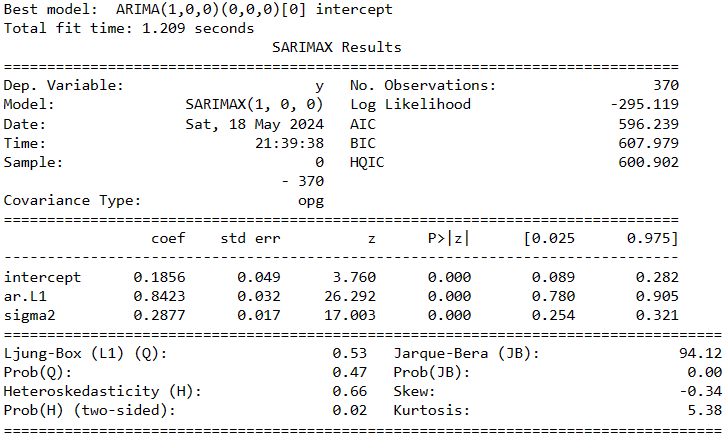




**Metrics:**

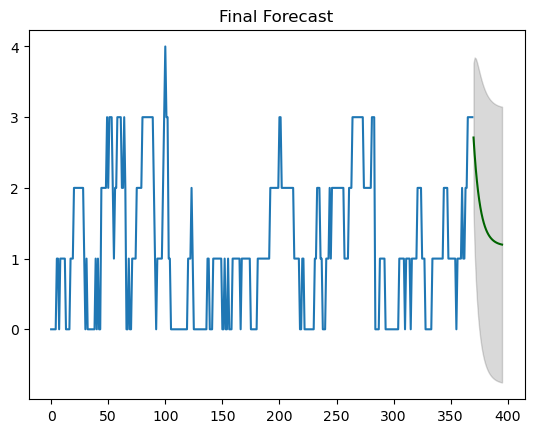
* 'mape': inf,
* 'me': -0.10454387928938709,
* 'mae': 0.5603780180862202,
* 'mpe': inf,
* 'rmse': 0.8001748337050231,
* 'corr': 0.08787424800675657,
* 'minmax': 0.3412174622343389

**Auto Arima:**

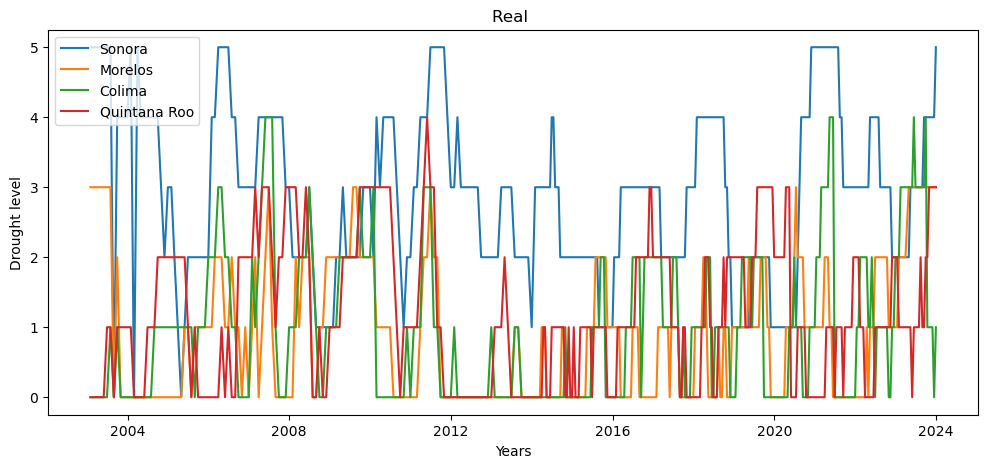
****

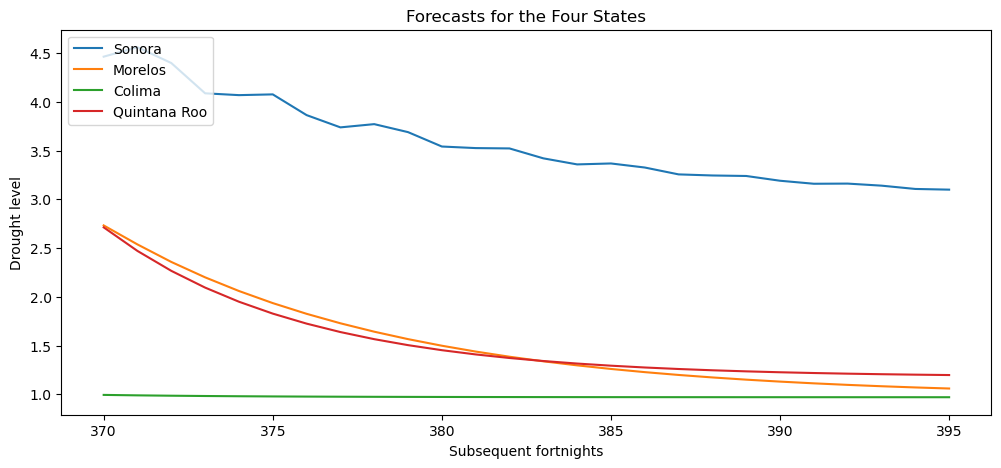
The AIC has reduced to 596 from 598[**¶**](#Auto-Arima)

**Forecast**:



# Results

In the graph below, we can see the 4 series of actual data.

We plotted the forecasts for the 4 data series to observe them together:

**Conclusion:**[**¶**](#Conclusion:)

The models are not functioning correctly to forecast the entire year. Despite using one model for each state, the results for all were very poor. They all show a downward trend, and their values are not distributed across the 5 levels; instead, they all tend to approach 1.

Sonora had the highest levels, and compared to the others, its forecast fluctuates between 2 levels. It is observed that it rises and falls slightly, but not as expected, since it remains within the same range. Morelos and Quintana Roo have very similar results, despite Quintana Roo having a higher mean (1.167, compared to Morelos' 0.89).

In comparison, Colima performed the worst, as its values never exceeded 1 and remained linear. As we have seen in its graph, Colima's values are typically distributed across the 4 drought levels.

Although various actions were taken to train our model with the best parameters, we need to further improve our data series and spend more time understanding the ARIMA model fully to comprehend what might be causing these results.

That is why the graphs and metrics we obtained were not discussed in detail, as they all would have shown incorrect results due to our model being poorly trained

This will be addressed as future actions.

# References

Amat Rodrigo, J. and Escobar Ortiz, J. (2023). *ARIMA and SARIMAX models with python*. [online] cienciadedatos.net. Available at: https://cienciadedatos.net/documentos/py51-arima-sarimax-models-python [Accessed 18 May 2024].

Brownlee, J. (2017a). *A Gentle Introduction to Autocorrelation and Partial Autocorrelation*. [online] Machine Learning Mastery. Available at: https://machinelearningmastery.com/gentle-introduction-autocorrelation-partial-autocorrelation/.

Epicentro Geografico. (2021). *Ríos de México: vertientes y cuencas - Epicentro Geográfico*. [online] Available at: https://epicentrogeografico.catedrauno.com/2021/02/rios-de-mexico-vertientes-cuencas-mapas/ [Accessed 28 Mar. 2024].

Estadística, P. y (2024). *▷ Modelo ARIMA*. [online] Probabilidad y Estadística. Available at: https://www.probabilidadyestadistica.net/modelo-arima/ [Accessed 19 May 2024].

Frost, J. (2021). *Autocorrelation and Partial Autocorrelation in Time Series Data*. [online] Statistics By Jim. Available at: https://statisticsbyjim.com/time-series/autocorrelation-partial-autocorrelation/.

Geografía (INEGI), I.N. de E. y (n.d.). *Quiénes somos*. [online] en.www.inegi.org.mx. Available at: https://en.www.inegi.org.mx/inegi/quienes\_somos.html.

Jason Brownlee (2017b). *How to Create an ARIMA Model for Time Series Forecasting in Python*. [online] Machine Learning Mastery. Available at: https://machinelearningmastery.com/arima-for-time-series-forecasting-with-python/.

Los Mundos de México. (2019). *Regiones de México de México para explorar en nuestros viajes y aventuras.* [online] Available at: https://losmundosdemexico.com/regiones-demexico/.

Nota informativa Situación del agua en México. (2023). [online] *IMCO.org*. Instituto Mexicano para la Competitividad A.C. Available at: https://imco.org.mx/wp-content/uploads/2023/02/Nota-IMCO\_La-situacion-del-agua-en-Mexico-1.pdf#:~:text=M%C3%A9xico%20debe%20abordar%20la%20problem%C3%A1tica%20del%20agua%20con [Accessed 28 Mar. 2024].

Paialunga, P. (2022). *Weather forecasting with Machine Learning, using Python*. [online] Medium. Available at: https://towardsdatascience.com/weather-forecasting-with-machine-learning-using-python-55e90c346647.

Prabhakaran, S. (2019a). *KPSS Test for Stationarity*. [online] Machine Learning Plus. Available at: https://www.machinelearningplus.com/time-series/kpss-test-for-stationarity/#:~:text=KPSS%20test%20is%20a%20statistical%20test%20to%20check [Accessed 19 May 2024].

Prabhakaran, S. (2019b). *Time Series Analysis in Python - A Comprehensive Guide with Examples – ML+*. [online] Machine Learning Plus. Available at: https://www.machinelearningplus.com/time-series/time-series-analysis-python/.

Prabhakaran, S. (2021). *ARIMA Model - Complete Guide to Time Series Forecasting in Python | ML+*. [online] Machine Learning Plus. Available at: https://www.machinelearningplus.com/time-series/arima-model-time-series-forecasting-python/?utm\_content=cmp-true.

pythonbasics.org. (n.d.). *Seaborn Distplot - Python Tutorial*. [online] Available at: https://pythonbasics.org/seaborn-distplot/.

Salvador, D.G.C., Ernesto Núñez Flores, Roberto Ramírez García, Julio César Romero (2021). *La sequía severa en México*. [online] Ciencia UNAM. Available at: https://ciencia.unam.mx/leer/1146/la-sequia-severa-en-mexico [Accessed 28 Mar. 2024].

Secretaría de Medio Ambiente y Recursos Naturales (2021). *Estadísticas del Agua en México*. [online] Available at: https://agua.org.mx/wp-content/uploads/2023/10/EAM\_2021.pdf.

smn.conagua.gob.mx. (2024). *Monitor de Sequía en México*. [online] Available at: https://smn.conagua.gob.mx/es/climatologia/monitor-de-sequia/monitor-de-sequia-en-mexico [Accessed 27 Mar. 2024].

Stephanie (2016). *KPSS Test: Definition and Interpretation*. [online] Statistics How To. Available at: https://www.statisticshowto.com/kpss-test/.

Torres, A. (2024). *Mexico City is running out of water due to a drought crisis*. [online] Mail Online. Available at: https://www.dailymail.co.uk/news/article-13215895/Mexico-City-running-water-drought-crisis-day-zero.html [Accessed 28 Mar. 2024].

www.gob.mx. (n.d.). *Comisión Nacional del Agua | Gobierno* . [online] Available at: https://www.gob.mx/conagua/que-hacemos.