Weather Pattern Prediction of the Four Corner States

COMP 4441, Summer 2021

Benjamin Karabinus
Marcelo Ellmann Clemente

Research Question

Background

In the past decades, there has been a noticeable increase in the volatility of weather patterns in the Western United States.

Where are we looking at?

In particular, the four corners region (Arizona, Colorado, New Mexico and Utah) have experienced prolonged drought, monsoon rainfall, and a noticeable increase in the intensity and duration of heat waves.

What will be done?

To better understand seasonal weather patterns, this report will present a univariate series analysis of mean monthly temperature, total monthly precipitation and snowfall, in the period between January 1970 and December 2019.

Research Question Deep-Dive

Question 1

Is there a noticeable trend in monthly mean temperature, total monthly precipitation and snowfall?

Question 2

Is the SARIMA model an effective means of visualizing and predicting mean monthly temperature, total monthly precipitation and total monthly snowfall?

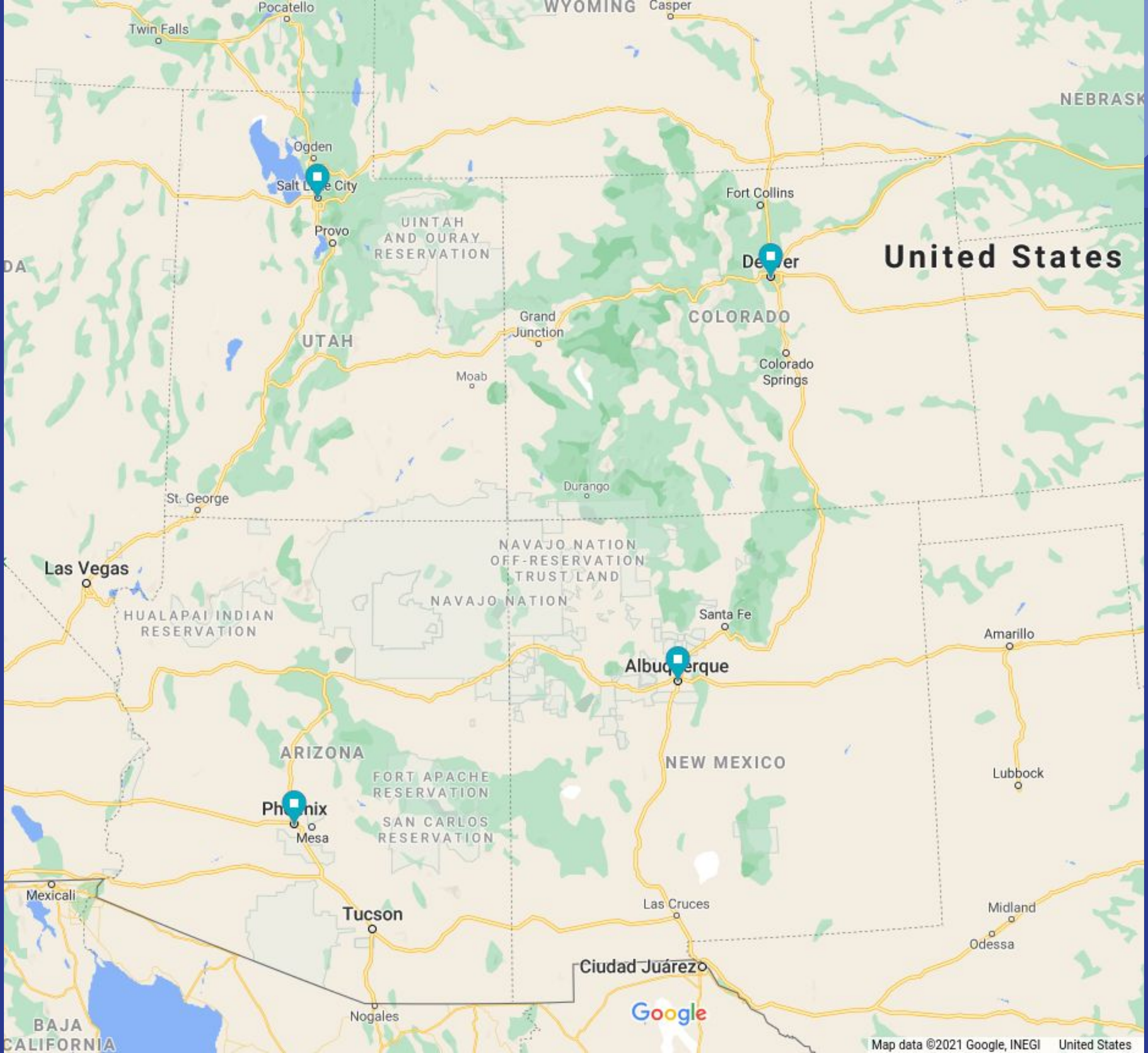
Time-series analysis

The time-series analysis of the weather pattern of the four-corner states will be carried using a SARIMA model (Seasonal Autoregressive Integrated Moving Average).

An extension of ARIMA, which does not support seasonal data, i.e. a time-series with a repeating cycle (e.g. weather, air-travel passengers, sports tickets sales, etc.).

The capitals of the Four-Corner States were chosen as they represent the biggest population centers in each of said states, making them the regions where the weather volatility patterns have the greatest impact.

Their geographical location, being disperse, provides diverse weather patterns that our model can be executed upon.



Pocatello
Twin Falls
WYOMING
Casper
Ogden
Provo
UINTAH AND OURAY RESERVATION
Fort Collins
United States
UTAH
Grand Junction
COLORADO
Colorado Springs
Moab
Durango
St. George
NAVAJO NATION OFF-RESERVATION TRUST LAND
NAVAJO NATION
Las Vegas
HUALAPAI INDIAN RESERVATION
Santa Fe
Amarillo
ARIZONA
FORT APACHE RESERVATION
NEW MEXICO
Lubbock
Mesa
SAN CARLOS RESERVATION
Mexicali
Tucson
Las Cruces
Midland
Odessa
Ciudad Juárez
Nogales
BAJA CALIFORNIA
Google
Map data ©2021 Google, INEGI
United States

SARIMA?

A SARIMA model is the seasonal flavor of the autoregressive integrated moving average (ARIMA) model. ARIMA and SARIMA models are a statistical technique used to analyze times series data and in certain cases predict future values. The goal of autoregressive models is to predict future values of the target variable by regressing against past observations of the variable.

Seasonal autoregressive models are used to perform similar analysis while accounting for any seasonality observed in the time series being analyzed. ARIMA and SARIMA models also incorporate differencing to correct for data that is not stationary and a moving average component that includes errors in previous predictions as a parameter in the model.

Datasets Used

- NOAA (National Oceanic and Atmospheric Administration) through the Global Historical Climatology Network.
- The data is a composite of climate records from numerous sources that were merged and then subjected to a suite of quality assurance reviews. The archive includes over 40 meteorological elements including temperature daily maximum/minimum, temperature at observation time, precipitation, snowfall, snow depth, evaporation, wind movement, wind maximums, soil temperature, cloudiness, and more.

Datasets Continued

- Not all weather stations collect all data used in this study. For that reason, the weather station located within each city's major airport will be used as they provide the richest source of data for our study.
- Denver will be an exception as its major airport changed during the mid 90's from Stapleton to DIA. In this case we'll be using the weather data from Denver's Central Park.
- Our study will use the following measurements from January 1st 1970 throughout December 31st 2019 (50 years):
    - TAVG: daily average temperature in °F (ceiling((TMAX + TMIN)/2)), as a monthly average.
    - SNOW: daily snowfall precipitation in inches, as a monthly total.
    - PRCP: daily precipitation in inches, as a monthly total.

STATION	NAME	DATE	PRCP	SNOW	TAVG	TMAX	TMIN
USW00023062	DENVER CENTRAL PARK, CO US	1970-01-01	0.00	0.0	19	28	10
USW00023062	DENVER CENTRAL PARK, CO US	1970-01-02	0.00	0.0	20	30	10
USW00023062	DENVER CENTRAL PARK, CO US	1970-01-03	0.00	0.0	16	30	3
USW00023062	DENVER CENTRAL PARK, CO US	1970-01-04	0.02	0.3	17	33	1
USW00023062	DENVER CENTRAL PARK, CO US	1970-01-05	0.04	0.6	4	14	-6
USW00023062	DENVER CENTRAL PARK, CO US	1970-01-06	0.00	0.0	6	22	-10
USW00023062	DENVER CENTRAL PARK, CO US	1970-01-07	0.00	0.0	9	23	-5
USW00023062	DENVER CENTRAL PARK, CO US	1970-01-08	0.00	0.0	7	22	-8
USW00023062	DENVER CENTRAL PARK, CO US	1970-01-09	0.00	0.0	20	38	1
USW00023062	DENVER CENTRAL PARK, CO US	1970-01-10	0.00	0.0	38	52	23
USW00023062	DENVER CENTRAL PARK, CO US	1970-01-11	0.00	0.0	36	47	25
USW00023062	DENVER CENTRAL PARK, CO US	1970-01-12	0.00	0.0	36	52	21
USW00023062	DENVER CENTRAL PARK, CO US	1970-01-13	0.00	0.0	38	51	24
USW00023062	DENVER CENTRAL PARK, CO US	1970-01-14	0.00	0.0	40	55	25
USW00023062	DENVER CENTRAL PARK, CO US	1970-01-15	0.00	0.0	39	50	28
USW00023062	DENVER CENTRAL PARK, CO US	1970-01-16	0.00	0.0	28	37	20
USW00023062	DENVER CENTRAL PARK, CO US	1970-01-17	0.00	0.0	25	32	18
USW00023062	DENVER CENTRAL PARK, CO US	1970-01-18	0.00	0.0	36	53	20

Data Preparation

- Some weather stations have few (~1%) sparsely missing measurements of snow data, these are replaced with the median measurements of snowfall of each given city.
- Many TAVG measurements are missing from the datasets, but are easily calculated by ceiling((TMAX+TMIN)/2), which is consistent with the existing measurements.

```
ABQ$SNOW[is.na(ABQ$SNOW)] <- median(ABQ$SNOW, na.rm = T)
ABQTEMP <- data.frame(ABQ$TMAX, ABQ$TMIN)
ind <- which(is.na(ABQ), arr.ind=TRUE)
ABQ[ind] <- round(rowMeans(ABQTEMP, na.rm=TRUE)[ind[,1]],0)
```

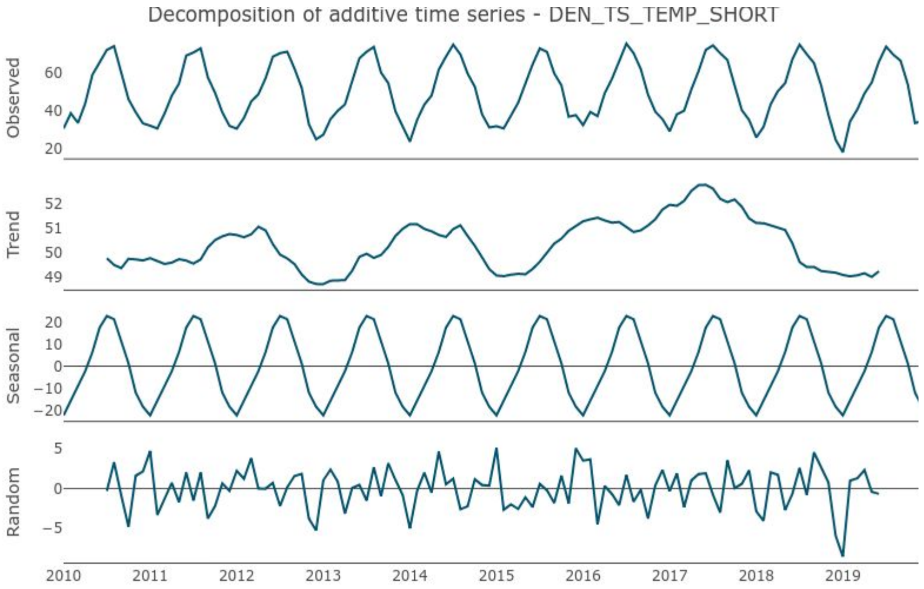
Testing the time series

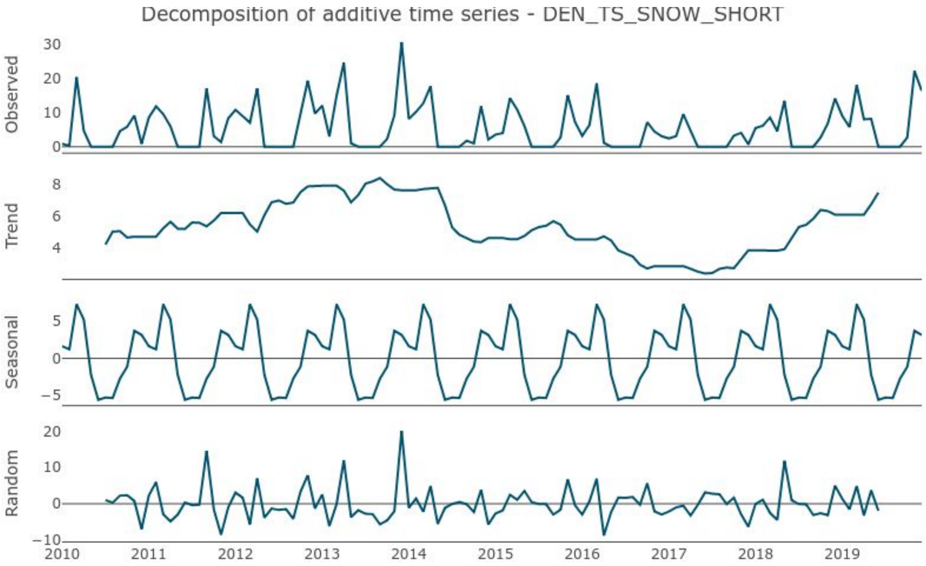
- In order to apply the SARIMA model we need to test whether the time series are stationary or not.
- Stationary time series is when the mean and variance are constant over time. It is easier to make predictions when the series are stationary.

```
> adf.test(ABQ_TS_TAVG,alternative = "stationary")

	Augmented Dickey-Fuller Test

data:  ABQ_TS_TAVG
Dickey-Fuller = -15.783, Lag order = 8, p-value = 0.01
alternative hypothesis: stationary
```

- Given the ADF test results, we have evidence to reject the null-hypothesis and assume that our time series is stationary, so we can apply the SARIMA model.

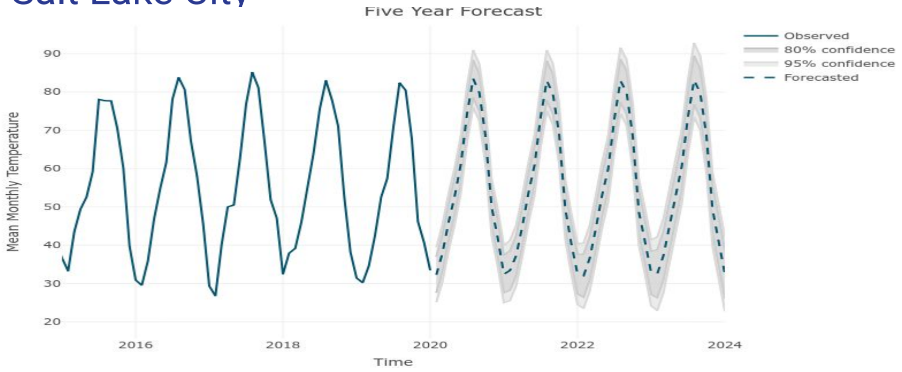
Seasonal/Trend/Remainder Components Graphs



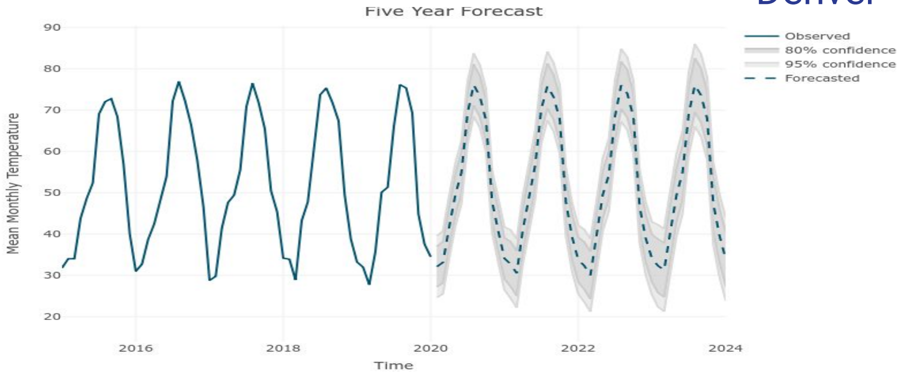


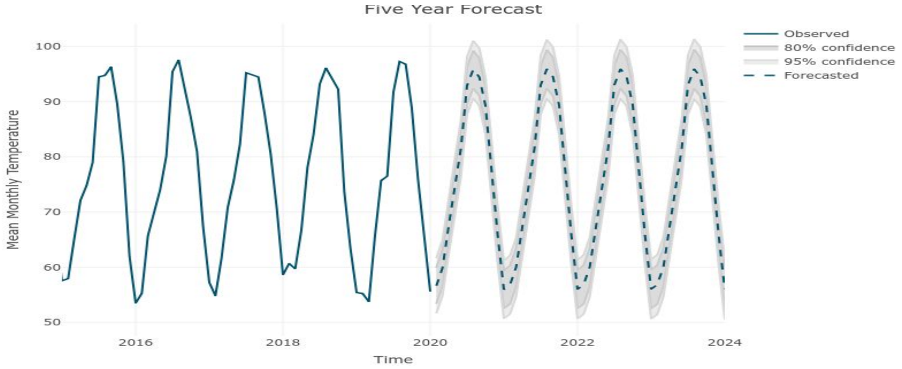
Five Year Average Temperature Forecast

Salt Lake City

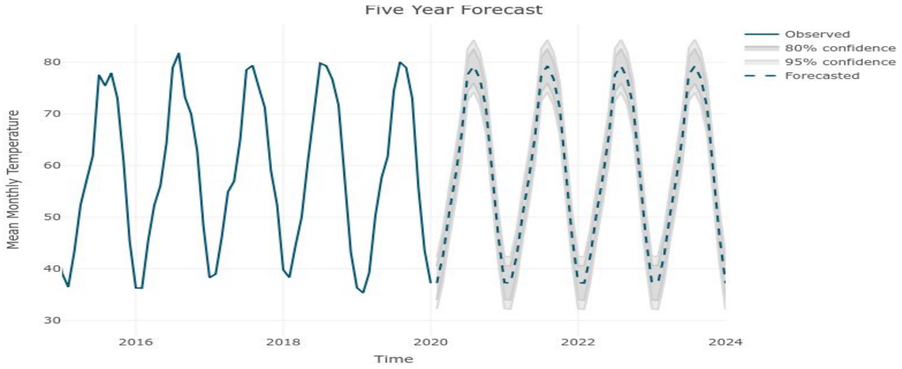


Denver





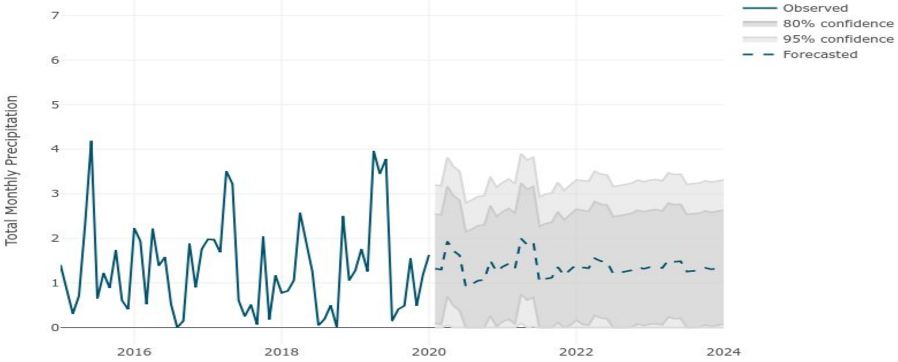
Phoenix



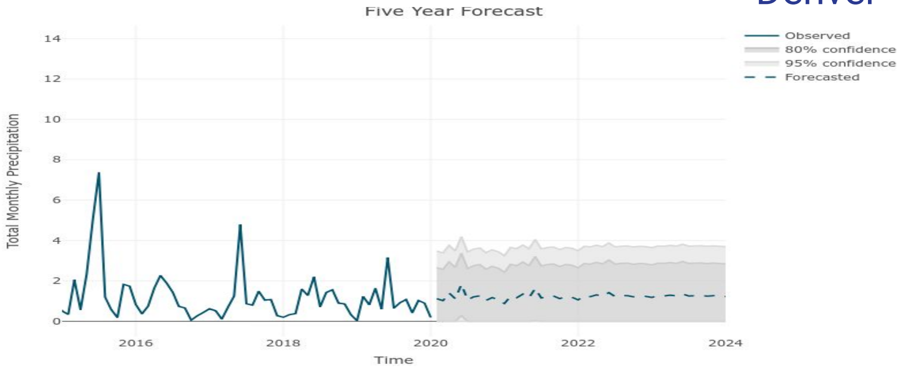
Albuquerque

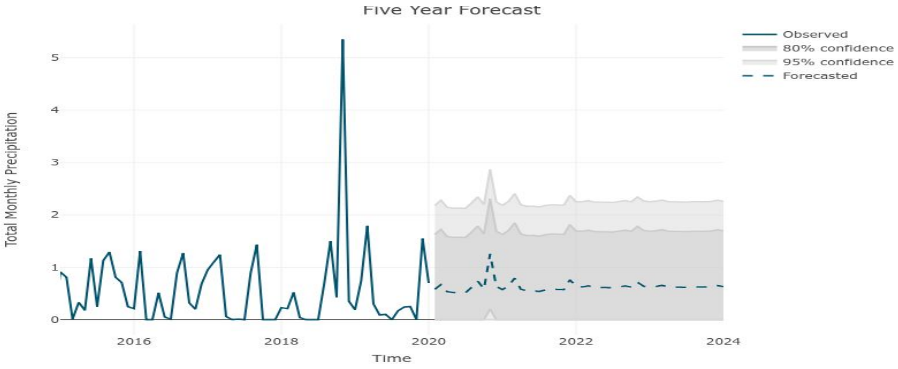
Five Year Average Precipitation Forecast

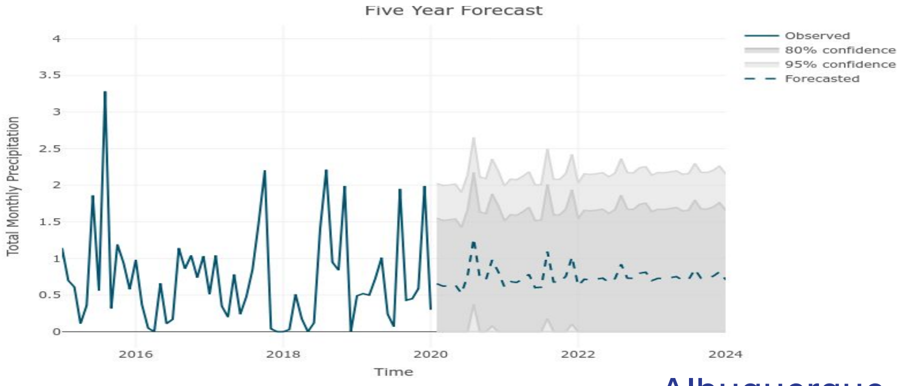
Salt Lake City



Denver





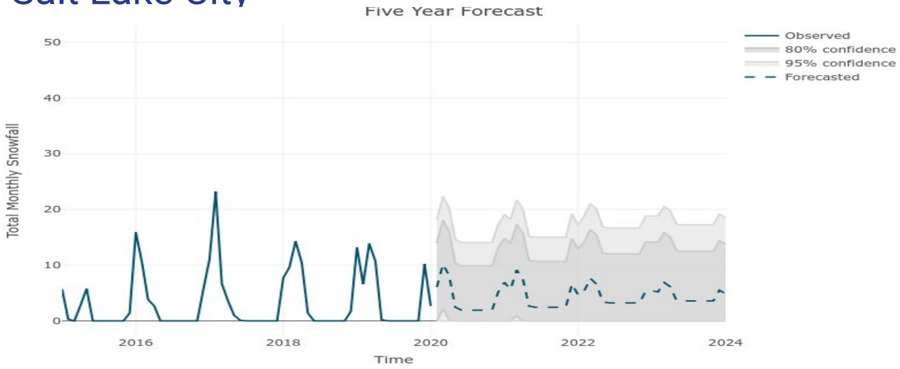


Phoenix

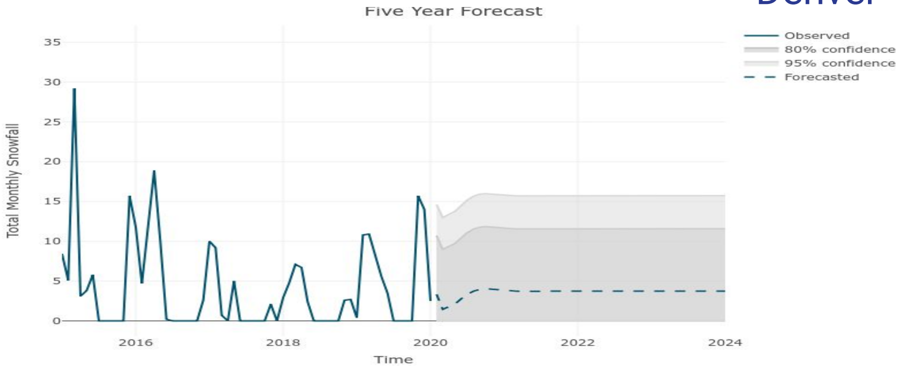
Albuquerque

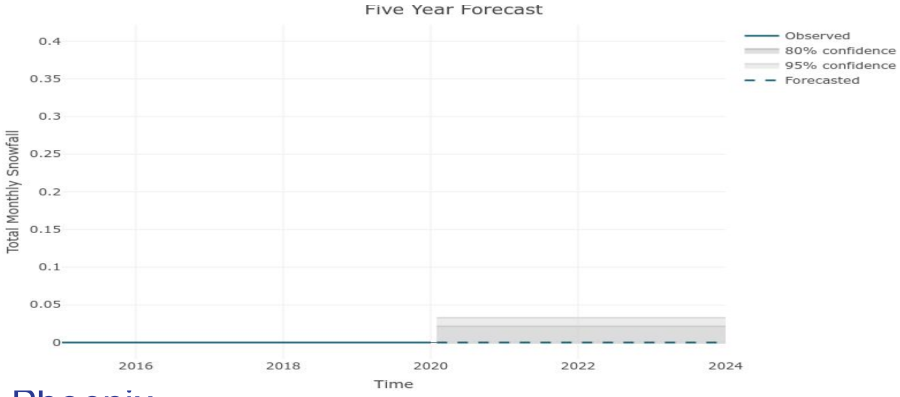
Five Year Average Snowfall Forecast

Salt Lake City

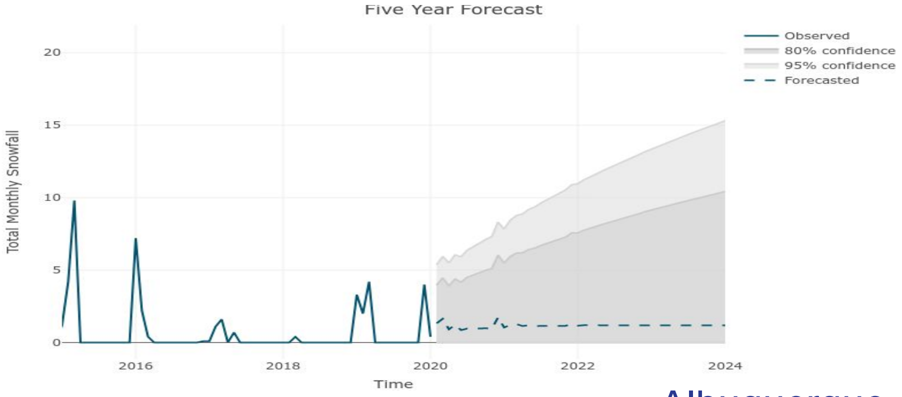


Denver





Phoenix



Albuquerque

Conclusion

Question 1

A slight increase in the average temperature in Albuquerque and Phoenix around 2016, no substantial changes or trends were noted in none of the average temperatures of the cities that were analyzed.

Question 2

Although SARIMA proved to work as an efficient method of predicting average temperatures, it was however, underwhelming in proving itself as a reliable method to predict total precipitation and snowfall.