Colombia_code

2025-04-02

Importing Data

```
## YEAR  JAN  FEB  MAR  APR  MAY  JUN  JUL  AUG  SEP  OCT  NOV  DEC  MAM  JJA  SON
##    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0
##  DJF  ANN
##    1    0

## YEAR  JAN  FEB  MAR  APR  MAY  JUN  JUL  AUG  SEP  OCT  NOV  DEC  MAM  JJA  SON
##    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0
##  DJF  ANN
##    0    0

## Rows: 124 Columns: 18
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## dbl (18): YEAR, JAN, FEB, MAR, APR, MAY, JUN, JUL, AUG, SEP, OCT, NOV, DEC, ...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.

## # A tibble: 6 x 18
##    YEAR   JAN   FEB   MAR   APR   MAY   JUN   JUL   AUG   SEP   OCT   NOV   DEC
##   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1  1901  90.1 102.   151.  209.  278.  256.  312.  329.  230.  288.  236.  128.
## 2  1902 138.   96.9  188.  230.  261.  254.  227.  226.  225   233.  198.  119.
## 3  1903  99.7  98.3  136.  230.  278.  336.  226.  316.  223.  226.  215.  148.
## 4  1904 111   112.   189.  271.  304   250.  264.  228.  219.  262.  169.  114.
## 5  1905 112.   87.7  142.  239.  286.  273.  248.  211.  273   257.  231.  163.
## 6  1906  90   108.   147.  271.  297.  309.  287.  238.  195.  252   192.  130.
## # i 5 more variables: MAM <dbl>, JJA <dbl>, SON <dbl>, DJF <dbl>, ANN <dbl>
```

Data Cleaning & TS Creation

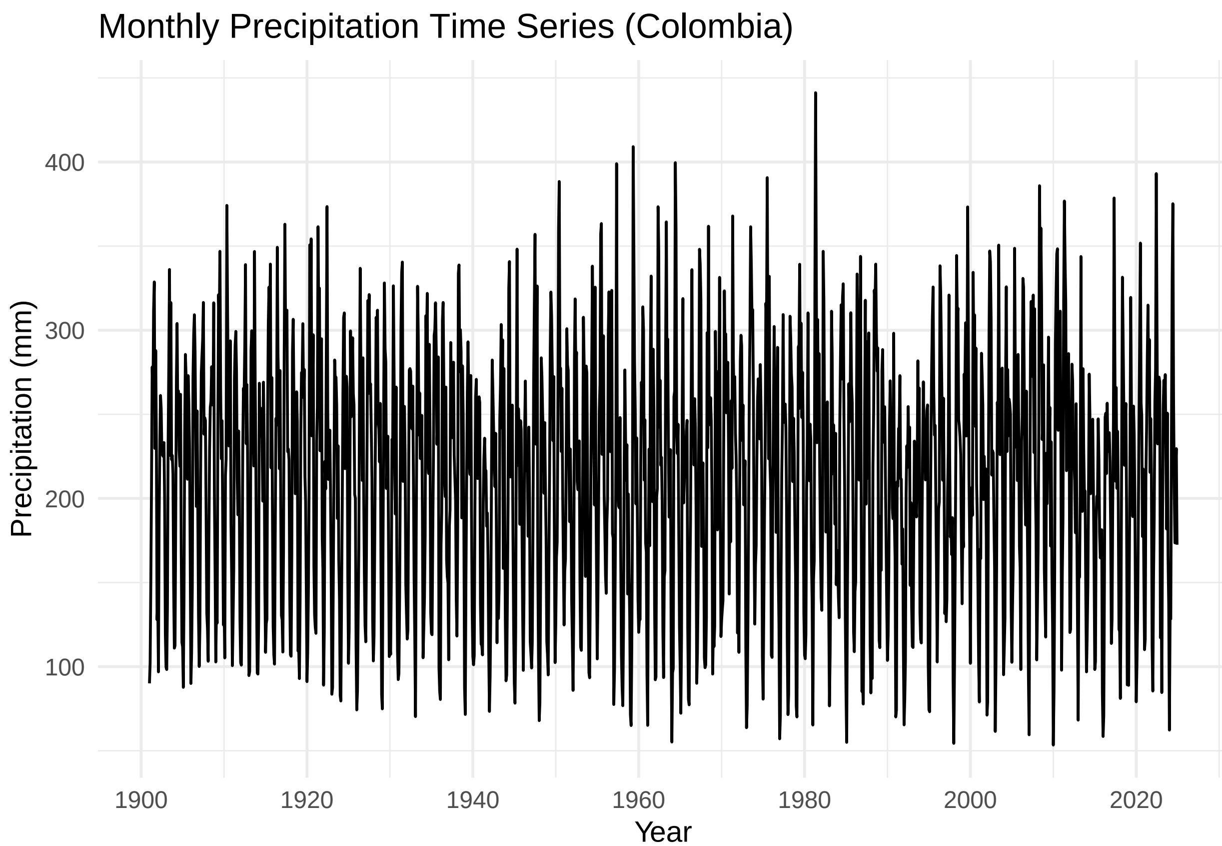
Table 1: First 10 Rows of Monthly Precipitation Data for Colombia

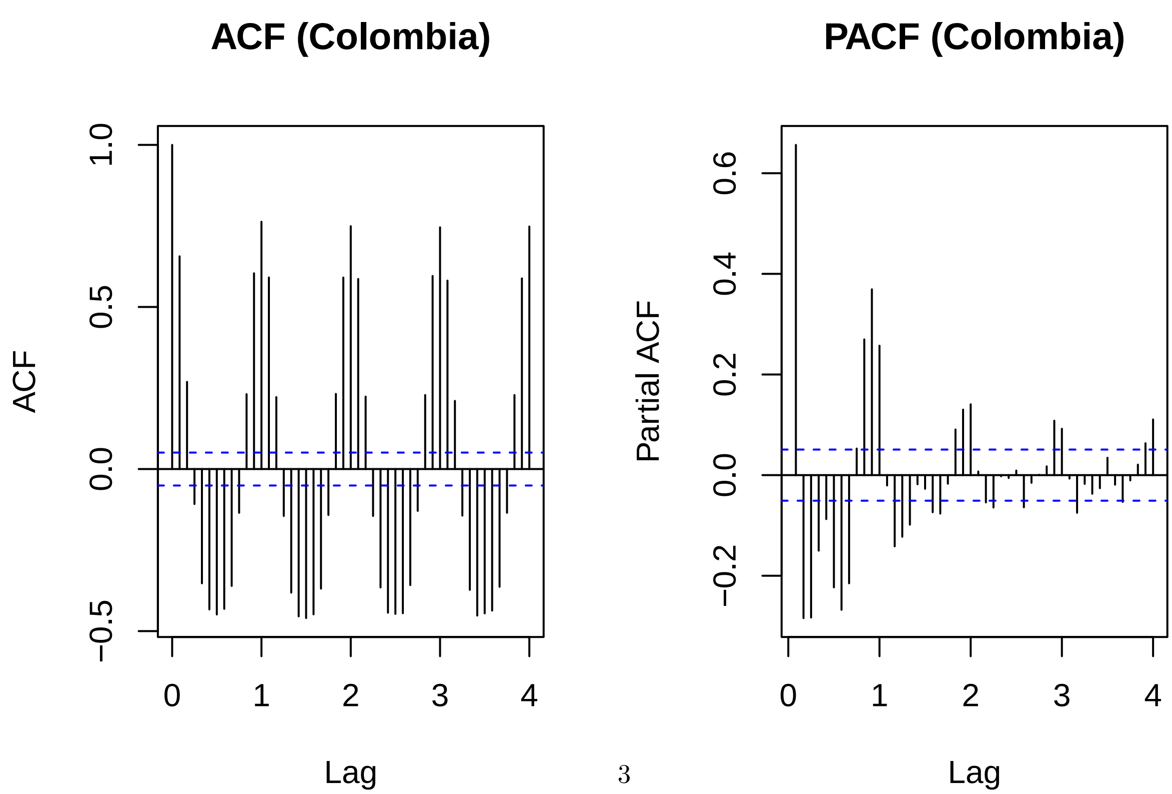
Month_Num	Date	YEAR	Month	Precipitation
1	1901-01-01	1901	JAN	90.1
2	1901-02-01	1901	FEB	102.2
3	1901-03-01	1901	MAR	150.7
4	1901-04-01	1901	APR	209.3
5	1901-05-01	1901	MAY	278.1

Month_Num	Date	YEAR	Month	Precipitation
6	1901-06-01	1901	JUN	255.6
7	1901-07-01	1901	JUL	311.9
8	1901-08-01	1901	AUG	328.7
9	1901-09-01	1901	SEP	229.7
10	1901-10-01	1901	OCT	287.9

Creating Time Series Object

Initial Plots



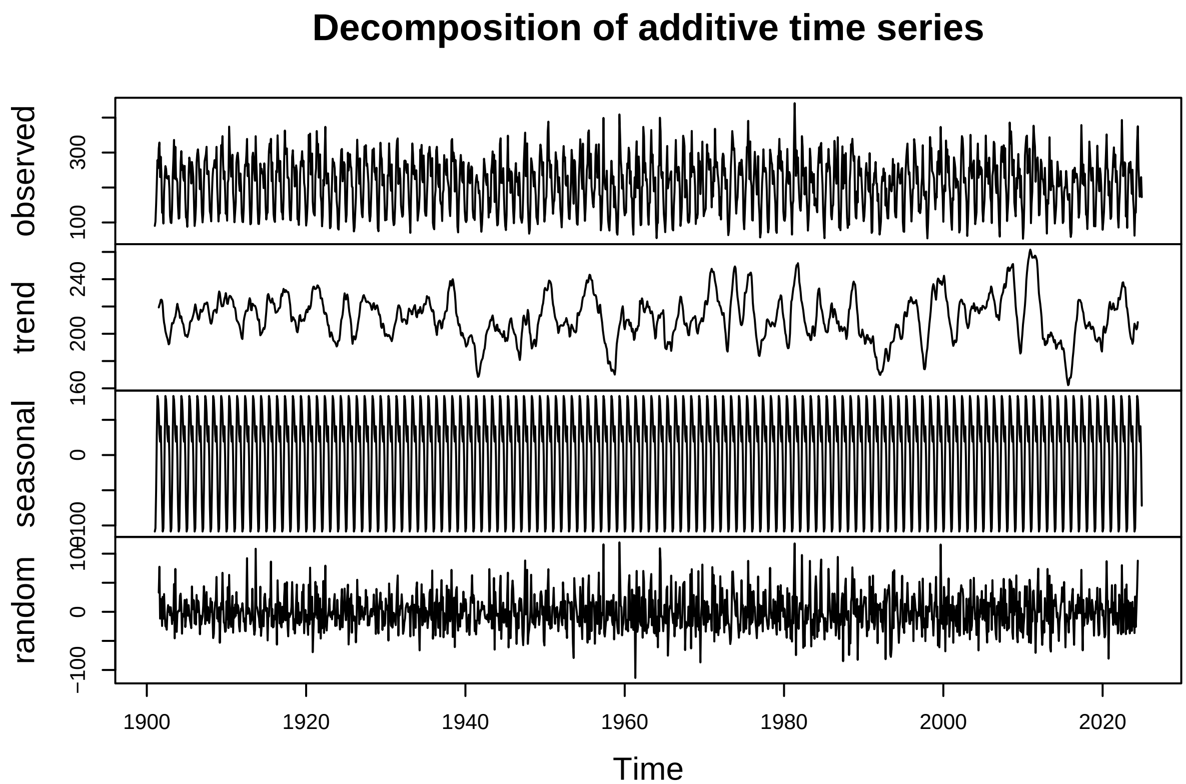


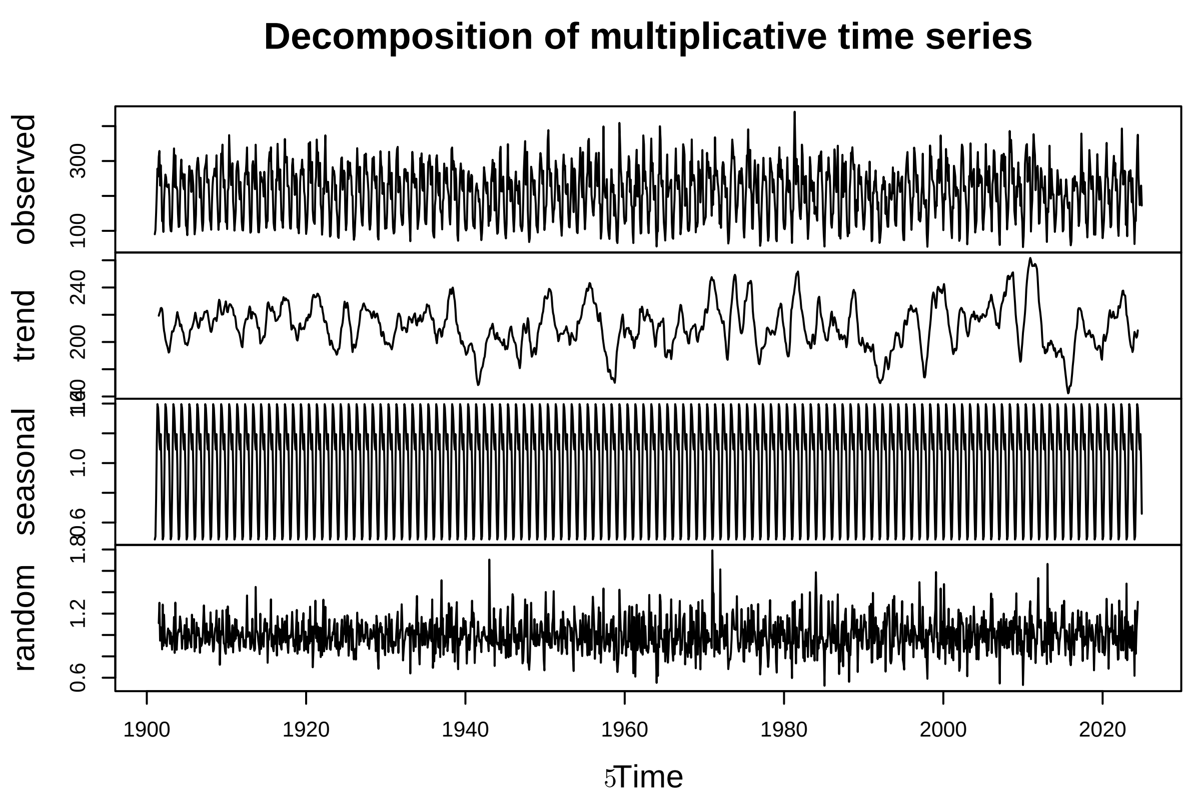
Interpretation: The time series graph seems to show seasonality as there are visible peaks and troughs, consistent with a seasonal climate.There's no clear upward or downward trend, suggesting the mean precipitation level has remained relatively stable over time. Moreover, although there's variation in the amplitude showing more extreme years, this doesn't appear to be systematically increasing or decreasing.

The ACF shows a wave like pattern, and the significant autocorrelation at multiple repetitive lags suggests strong seasonality, likely due to Colombia's bi-modal rainy seasons. The PACF has a significant spike at lag 1 and smaller spikes at subsequent lags, suggesting a short-term auto regressive (AR) component, where current precipitation is influenced by the previous month or two.

Testing and Training

Decomposing Time Series





Interpretation: Additive Model - the trend shows some variability across the years, but there's no clear increasing or decreasing pattern over time, more just oscillation representing variability instead of a linear trend.The seasonal component is constant over time, showing strong seasonality over the years, which is logical given rain patterns. The residuals appear roughly stable, with some years having anomylous peaks and trough representing random occurrences, but the centered nature of the residual component demonstrates that the additive model is a relatively good fit.

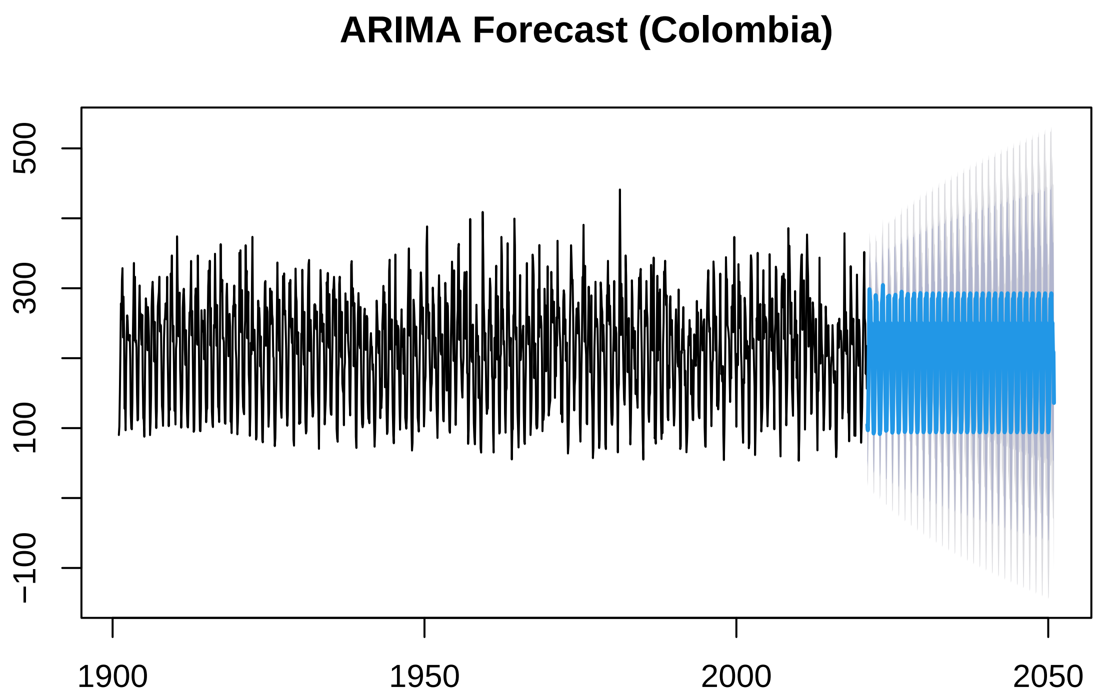
Multiplicative Model - The trend component is practically the same as in the additive model indicating that there is no increasing or decreasing linear trend.The seasonal component also looks the same, even though now the seasonal component is proportional to the trend, indicating no major changes. The residuals also appear relatively stable with some anomilies (more than in the additive model).

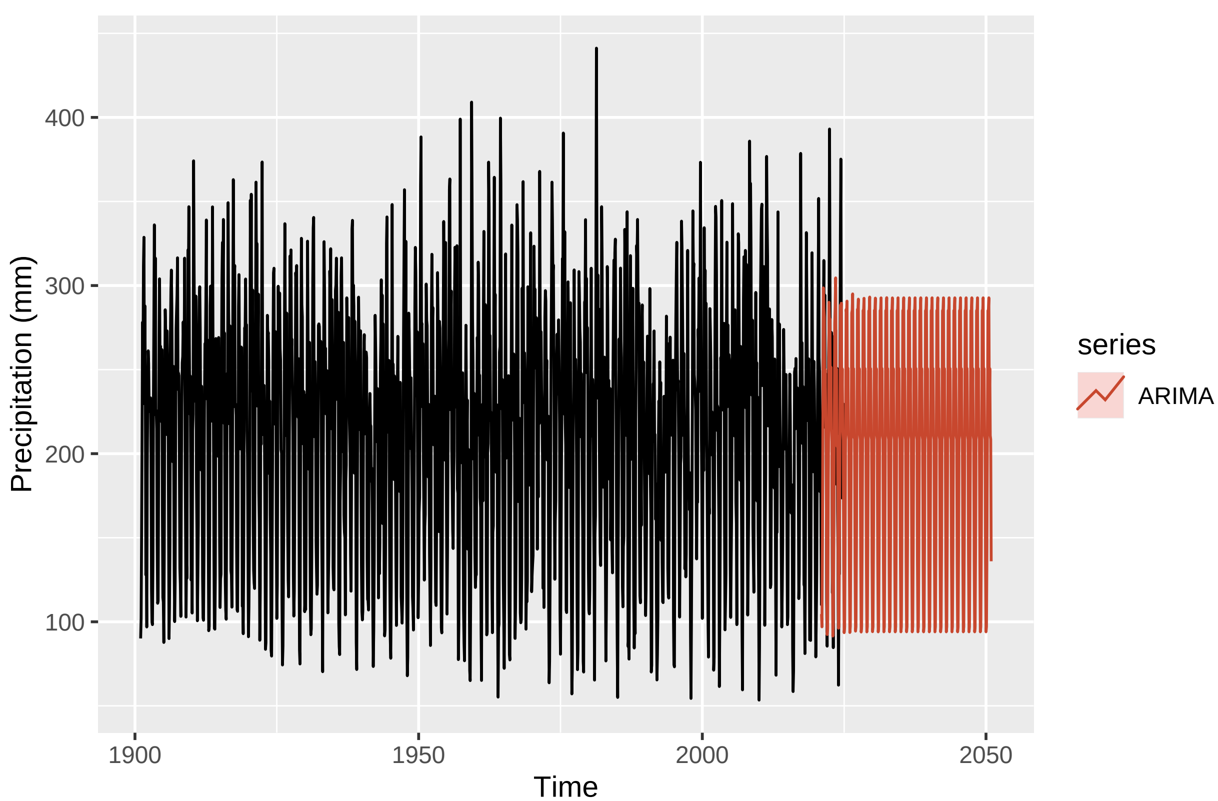
Detrending

```
##
## Call:
## lm(formula = c_monthly_data_long$Precipitation ~ t)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -157.793  -58.482    6.642   53.388  229.718
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.123e+02  3.855e+00  55.076   <2e-16 ***
## t           -8.393e-04  4.484e-03  -0.187    0.852
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 74.31 on 1486 degrees of freedom
## Multiple R-squared:  2.357e-05,  Adjusted R-squared:  -0.0006494
## F-statistic: 0.03503 on 1 and 1486 DF,  p-value: 0.8516
```

Interpretation: The slope is -0.000839mm/month of rain, indicating a very small but slight negative trend in precipitation. Nonetheless, the p-value is 0.852, which is considerably higher than 0.05, meaning that it's not statistically significant. The t value of 0.187 is also very close to 0 which indicates no meaningful relationship between monthly precipitation and time.

ARIMA


ARIMA Forecast (Colombia)
500
300
100
−100
1900
1950
2000
2050


Precipitation (mm)
400
300
200
100
1900
1950
2000
2050
Time
series
ARIMA

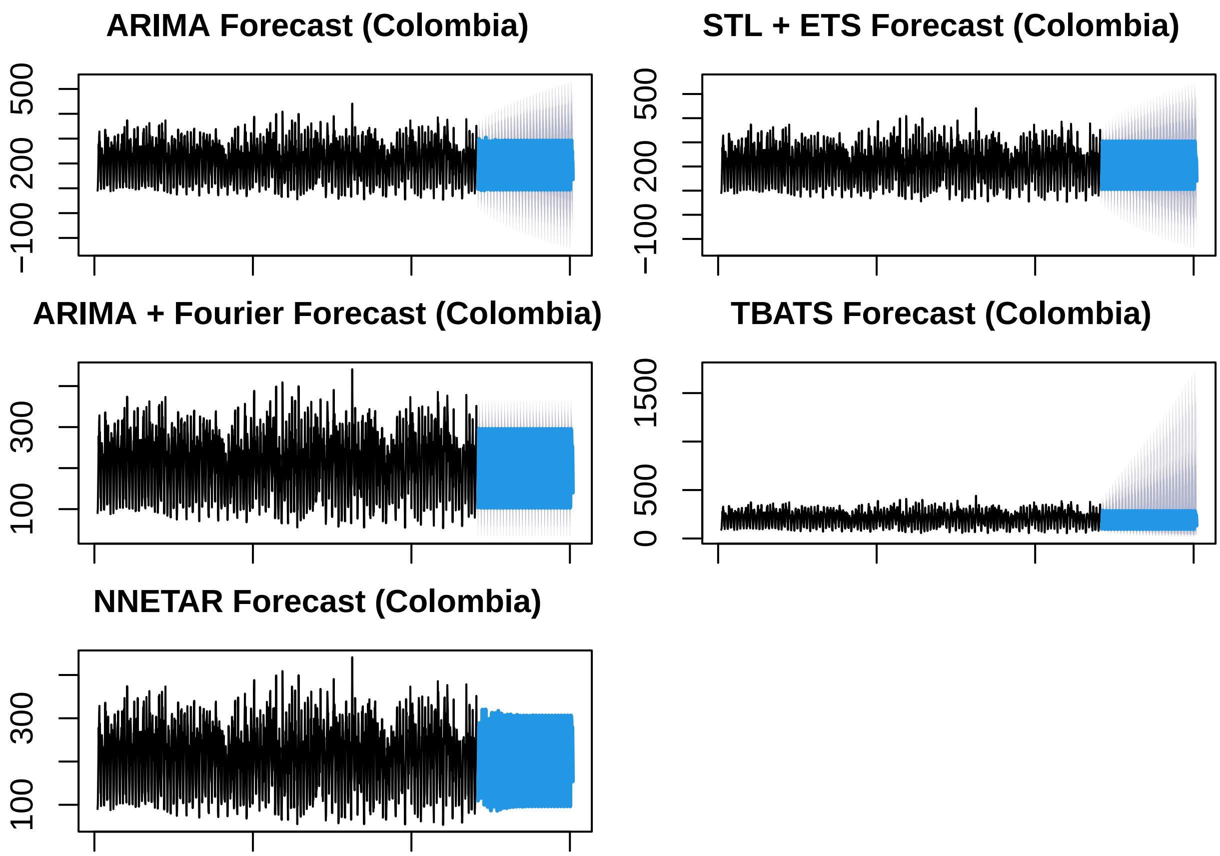
STL + ETS

ARIMA + Fourier terms

TBATS

Neural Network

Forecast Plots



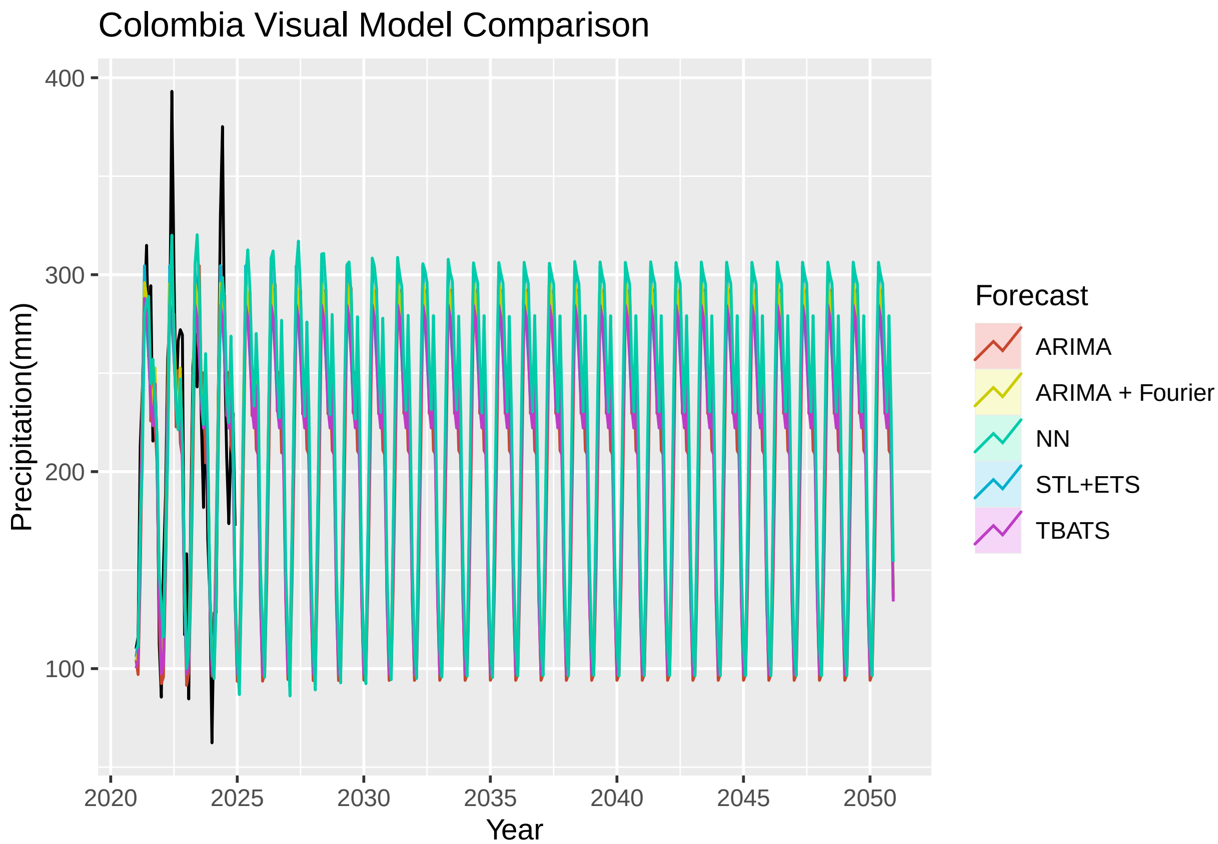
Model & Observed Data Plots

Scores

Table 2: Forecast Accuracy for Precipitation in Colombia

	ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's U
ARIMA	12.46980	42.04259	34.32337	3.32439	16.79040	-0.02992	0.62144
STL+ETS	5.73029	37.46420	29.10067	-1.20640	15.02950	0.01863	0.47917
ARIMA+Fourier	12.46980	42.04259	34.32337	3.32439	16.79040	-0.02992	0.62144
TBATS	12.60992	37.30779	28.73991	3.04859	14.42548	0.07198	0.53294
NN	3.77270	38.03449	32.31814	-1.72703	16.95058	0.14857	0.55210

Visual Model Comparison



Interpretation:

ARIMA: The ARIMA model captures seasonality relatively well, but the forecast band is wide, indicating increasing uncertainty over time. It has the lowest RMSE meaning the smallest magnitude of error and the best Theil U score, indicating that the model performs well compared to a naive model.

STL + ETS: The STL + ETS model shows a smooth forecast curve with seasonality retained. Although the variance increases, it remains relatively controlled. The model has the best MAPE, meaning that the average magnitude of the errors is the lowest, and the best residual autocorrelation, indicating that more of the errors are random and not correlated with previous error.

ARIMA + Fourier: This model is similar to regular ARIMA, but it captures cyclical seasonal components more clearly, which may not be so necessary here since the accuracy scores are the same as the ARIMA model.

TBATS: TBATS shows an aggressive increase in forecast uncertainty, which may indicate model over fitting or extrapolation issues. The forecast seems relatively unstable after 2030, and none of the accuracy scores outperform the other models, indicating that it isn't the best fit.

NNETAR: This model captures the general seasonal shape, but the forecast appears more rigid and less sensitive to variation. It is the worst performer across the models, with a high RMSE indicating a high magnitude of error, a high MAPE, and the worst Theil's U. Therefore, NNETAR isn't a good fit.