

# Climate Data Forecasting - Atmospheric $CO_2$ Concentration / Temperature / Precipitation

Wolfgang Vollmer

2026-01-08

## Contents

<b>1</b>	<b>Forecasting of Mauna Loa - Atmospheric Carbon Dioxide Analysis</b>	<b>2</b>
1.1	Stationarity and differencing . . . . .	2
1.1.1	Unitroot KPSS Test - fix number of seasonal differences/differences required . . .	3
1.1.2	Ljung-Box Test - independence/white noise of the time series . . . . .	3
1.1.3	ACF (Autocorrelation Function) Plots of Differences . . . . .	4
1.1.4	Time Series, ACF and PACF (Partial) Plots of Differences - for ARIMA p, q check	4
<b>2</b>	<b>ExponentiaL Smoothing (ETS) Forecasting Models</b>	<b>7</b>
2.1	ETS Models and their components . . . . .	8
2.1.1	Residual Accuracy with one-step-ahead fitted residuals - check RMSE, MAE . . .	9
2.1.2	Ljung-Box Test - independence/white noise of the forecasts residuals . . . . .	9
2.1.3	ETS Models - components of ETS(A,N,A), ETS(A,A,A), ETS(A,Ad,A), models .	10
2.1.4	Forecast Accuracy with Training/Test Data . . . . .	10
2.2	Forecasting with selected ETS model <ETS(A,A,A)> . . . . .	12
2.2.1	Forecast Plot of selected ETS model . . . . .	12
2.2.2	Residual Stationarity . . . . .	13
2.2.3	Ljung-Box Test and Histogram of forecast residuals with overlaid normal curve . .	14
<b>3</b>	<b>ARIMA Forecasting Models - AutoRegressive-Integrated Moving Average</b>	<b>15</b>
3.1	Seasonal ARIMA models . . . . .	15
3.1.1	Residual Accuracy with one-step-ahead fitted residuals - check RMSE, MAE . . .	15
3.1.2	Ljung-Box Test - independence/white noise of the forecasts residuals . . . . .	16
3.1.3	Forecast Accuracy with Training/Test Data . . . . .	16
3.2	CO2 - Forecasting with selected ARIMA model <ARIMA(1,1,1)(0,1,1)[12]> . . . . .	17
3.2.1	Forecast Plot of selected ARIMA model . . . . .	17
3.2.2	Residual Stationarity . . . . .	17
3.2.3	Ljung-Box Test and Histogram of forecast residuals with overlaid normal curve . .	18
<b>4</b>	<b>ARIMA vs ETS</b>	<b>20</b>
4.0.1	Comparing Residual and Forecast Accuracy of selected ETS and ARIMA model .	20
4.0.2	Forecast Plot of selected ETS and ARIMA model . . . . .	20
<b>5</b>	<b>Yearly Data Forecasts with ARIMA and ETS</b>	<b>23</b>
5.0.1	Comparing Residual and Forecast Accuracy of selected ETS and ARIMA model .	23
5.0.2	Forecast Plot of selected ETS and ARIMA model . . . . .	23

# 1 Forecasting of Mauna Loa - Atmospheric Carbon Dioxide Analysis

## 1.1 Stationarity and differencing

Stationary time series is one whose properties do not depend on the time at which the series is observed. Thus, time series with trends, or with seasonality, are not stationary — the trend and seasonality will affect the value of the time series at different times. On the other hand, a white noise series is stationary — it does not matter when you observe it, it should look much the same at any point in time.

Stationary time series will have no predictable patterns in the long-term. Time plots will show the series to be roughly horizontal (although some cyclic behaviour is possible), with constant variance.

If  $y_t$  is a *stationary* time series, then for all  $s$ , the distribution of  $(y_t, \dots, y_{t+s})$  does not depend on  $t$ .

If Time Series data with seasonality are non-stationary

- => first take a seasonal difference
- if seasonally differenced data appear are still non-stationary
- => take an additional first seasonal difference

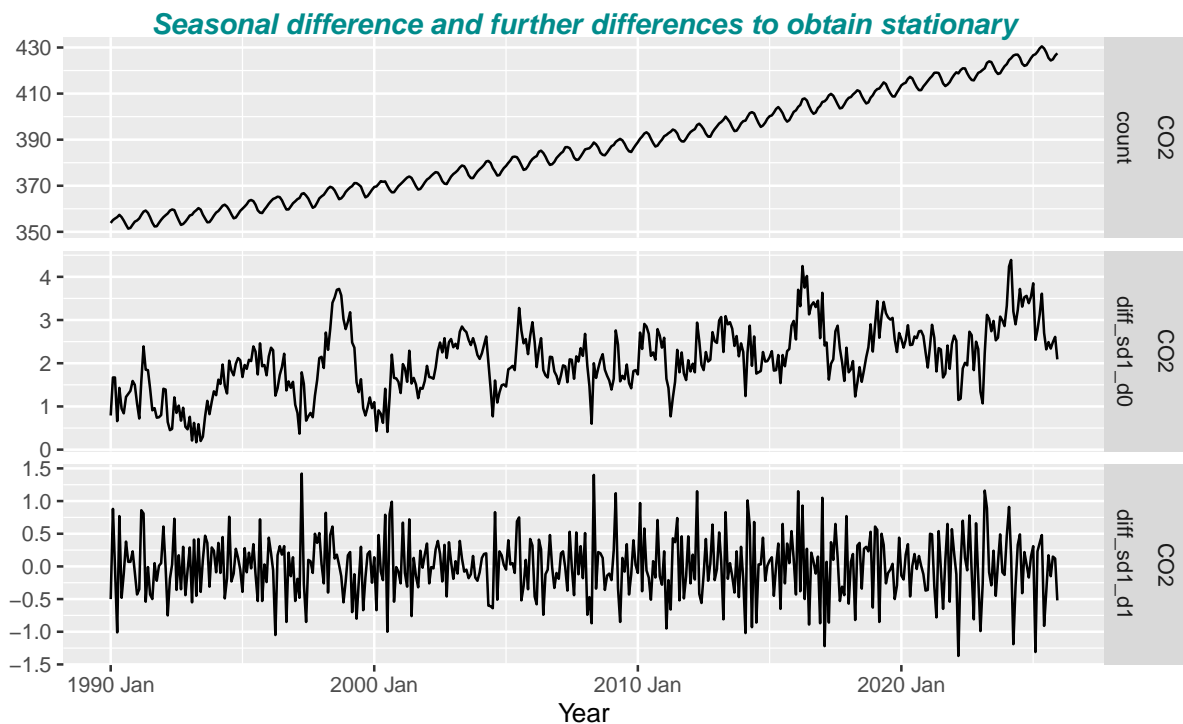
The model fit residuals have to be stationary. For good forecasting this has to be verified with residual diagnostics.

Essential:

- Residuals are uncorrelated
- The residuals have zero mean

Useful (but not necessary):

- The residuals have constant variance.
- The residuals are normally distributed.



### 1.1.1 Unitroot KPSS Test - fix number of seasonal differences/differences required

- For **ARIMA model** use Unitroot KPSS Test to fix number of differences
  - `unitroot_nsdiffs()` to determine  $D$  (the number of seasonal differences to use)
  - `unitroot_ndiffs()` to determine  $d$  (the number of ordinary differences to use)
  - The selection of the other model parameters ( $p, q, P$  and  $Q$ ) are all determined by minimizing the AICc
- kpss test of stationary of the differentiated data (with seasonal and ordinary differences) used in the **ARIMA model**
  - stationary times series: the distribution of  $(y_t, \dots, y_{t-s})$  does not depend on  $t$ .
  - *Null Hypothesis*  $H_0$ : stationary is given in the time series: data are stationary and non seasonal
    - \* for  $p < \alpha = 0.05$  : reject  $H_0$
    - \* for  $p >> \alpha = 0.05$  : conclude: differentiated data are stationary and non seasonal
- kpss test of stationary w/ `unitroot_nsdiffs` & `unitroot_ndiff` provides
  - minimum number of seasonal & ordinary differences required for a stationary series
  - first fix required seasonal differences and then apply `ndiffs` to the seasonally differenced data
  - returns 1 => for stationarity one seasonal difference resp. difference is required

```
#> ndiffs gives the number of differences required and
#> nsdifs gives the number of seasonal differences required
#> - to make a series stationary (test is based on the KPSS test
#> unitroot_kpss test to define seasonal (nsdifs) and ordinary (ndiffs) differences
#> # A tibble: 1 x 5
#>   Measure kpss_stat kpss_pvalue nsdifs ndiffs
#>   <fct>      <dbl>      <dbl>   <int>  <int>
#> 1 C02          11.4          0.01     1      1
#> unitroot_kpss test on (difference(count, 12) + difference())
#> # A tibble: 1 x 3
#>   Measure kpss_stat kpss_pvalue
#>   <fct>      <dbl>      <dbl>
#> 1 C02          0.00782      0.1
```

### 1.1.2 Ljung-Box Test - independence/white noise of the time series

The Ljung-Box Test becomes important when checking independence/white noise of the forecasts residuals of the fitted ETS resp. ARIMA models. There we have to check whether the forecast errors are normally distributed with mean zero

- `augment(fit) |> features(.innov, ljung_box, lag=x, dof=y)` (Ch. 5.4 Residual diagnostics)
  - portmanteau test suggesting that the residuals are white noise
  - *Null Hypothesis*  $H_0$ : independence/white noise for residuals, i.e. each autocorrelation in the time series residuals of the model for lag  $l$  is close to zero.
    - \* for  $p < \alpha = 0.05$  : reject  $H_0$
    - \* for  $p >> \alpha = 0.05$  : conclude: the residuals are not distinguishable from a white noise series
  - `lag = 2*m` (period of season, e.g.  $m=12$  for monthly season) | no season: `lag=10`
  - `dof = p + q + P + Q` (for ARIMA models only, degree of freedom)

Time series with trend and/or seasonality is not stationary. Tests are to be taken on the residuals of the fitted models.

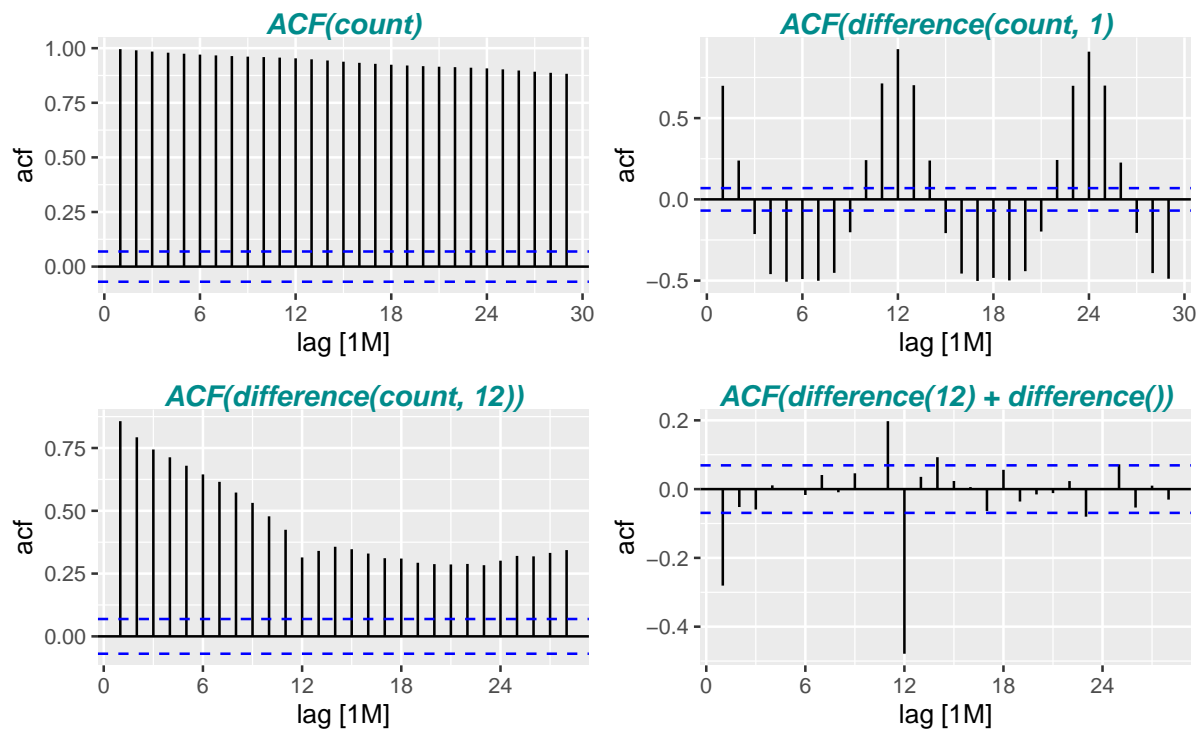
```
#> Ljung-Box test with (count), w/o differences
#> # A tibble: 1 x 3
#>   Measure lb_stat lb_pvalue
#>   <fct>      <dbl>      <dbl>
#> 1 C02          7709.          0
#> Ljung-Box test on (difference(count, 12))
#> # A tibble: 1 x 3
#>   Measure lb_stat lb_pvalue
#>   <fct>      <dbl>      <dbl>
```

```

#> 1 C02      3610.      0
#> Ljung-Box test on (difference(count, 12) + difference())
#> # A tibble: 1 x 3
#>   Measure lb_stat lb_pvalue
#>   <fct>      <dbl>      <dbl>
#> 1 C02      70.9  2.96e-11

```

### 1.1.3 ACF (Autocorrelation Function) Plots of Differences



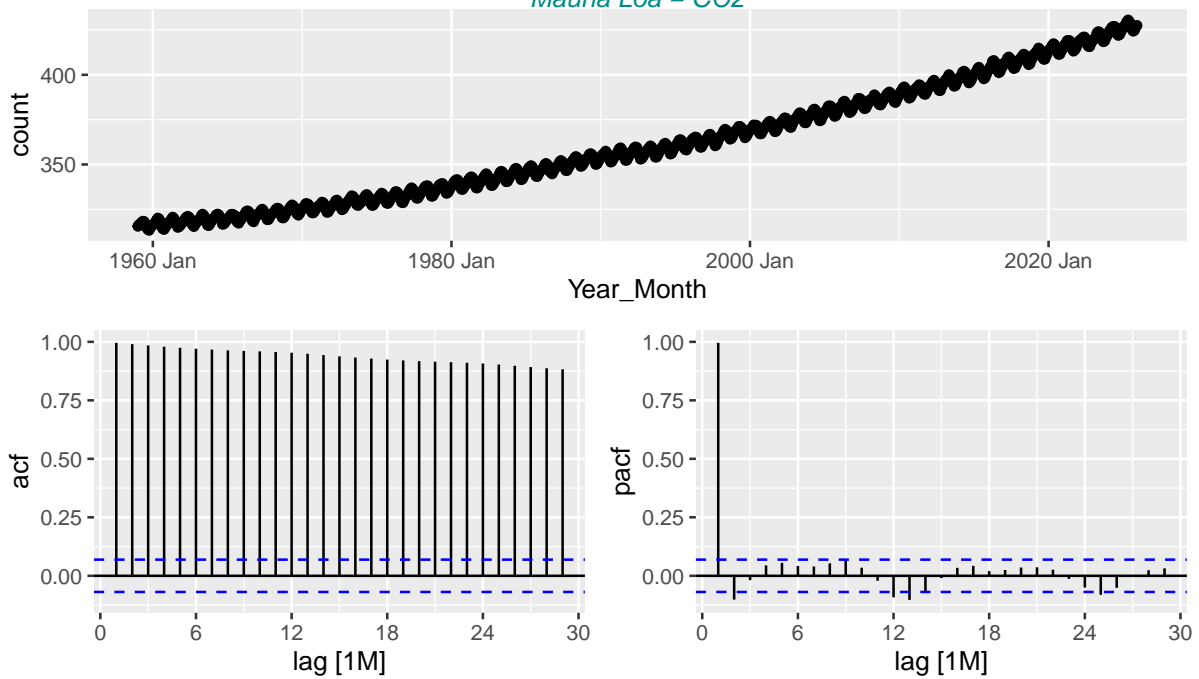
### 1.1.4 Time Series, ACF and PACF (Partial) Plots of Differences - for ARIMA p, q check

```

#> # A tibble: 1 x 4
#> # Groups:   City [1]
#>   City      Measure      Sum Mean
#>   <chr>    <fct>      <dbl> <dbl>
#> 1 Mauna Loa C02      290446. 361.

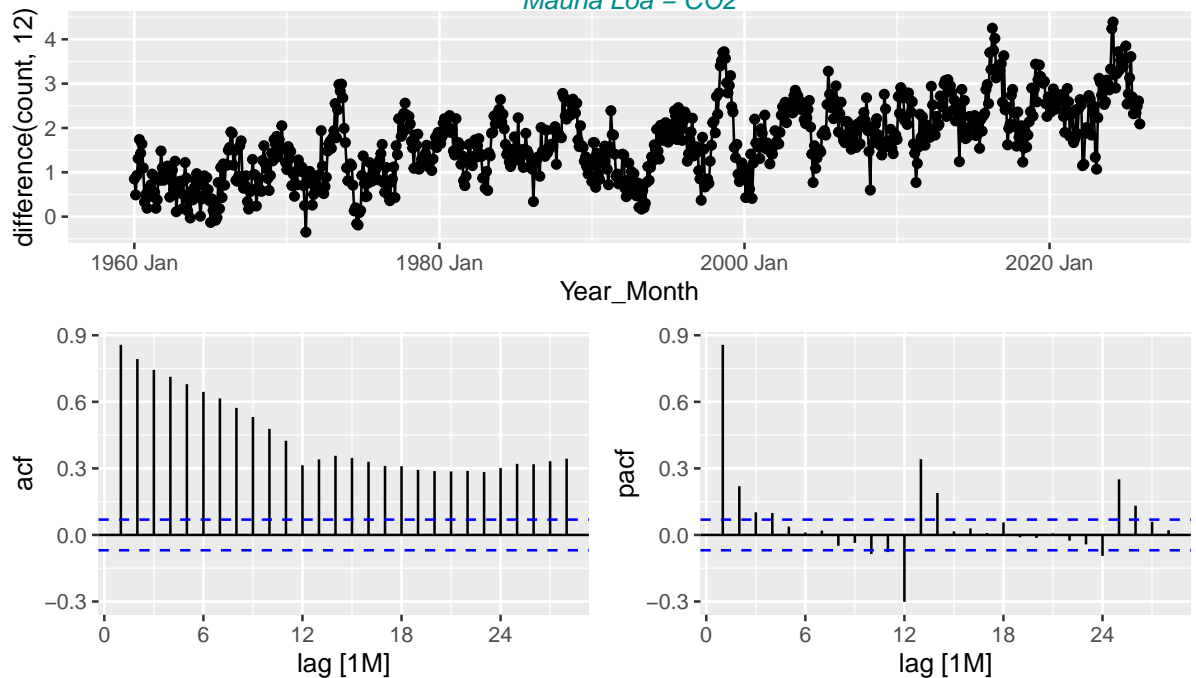
```

### Time Series, ACF & PACF for (count) Mauna Loa – CO2

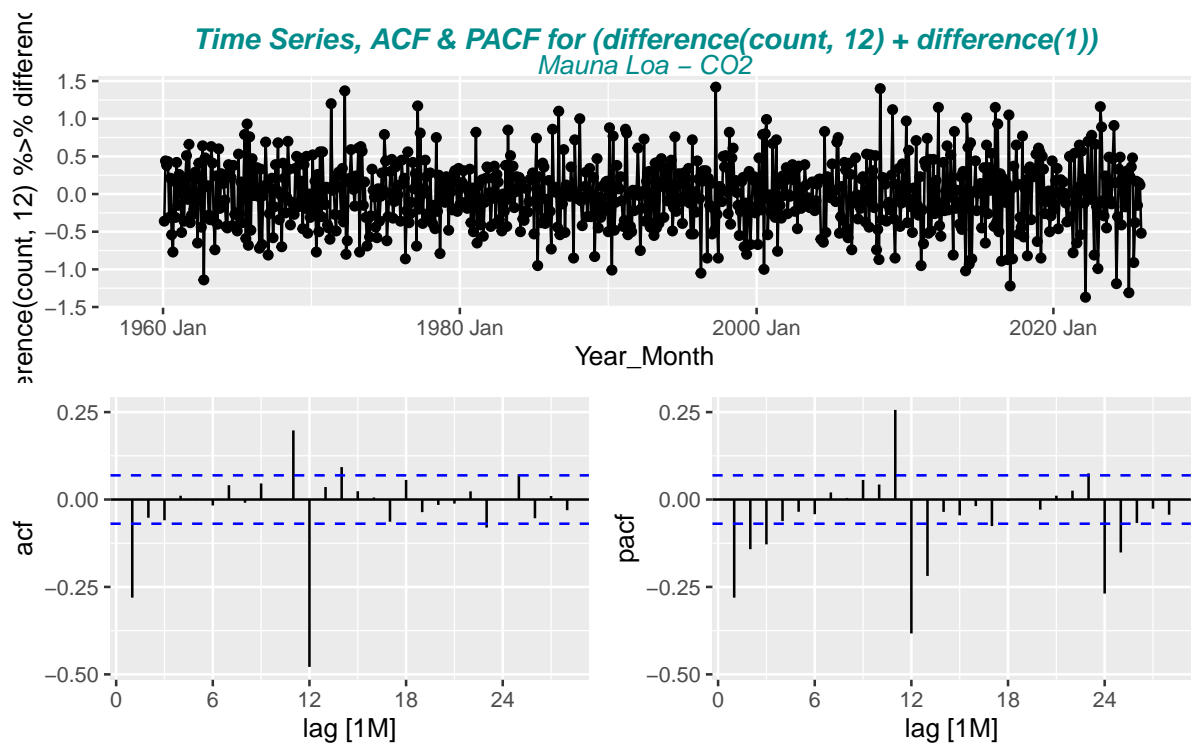


```
#> # A tibble: 1 x 4
#> # Groups:   City [1]
#>   City      Measure    Sum  Mean
#>   <chr>    <fct>    <dbl> <dbl>
#> 1 Mauna Loa CO2      1336.  1.69
```

### Time Series, ACF & PACF for (difference(count, 12)) Mauna Loa – CO2



```
#> # A tibble: 1 x 4
#> # Groups:   City [1]
#>   City      Measure    Sum  Mean
#>   <chr>    <fct>    <dbl> <dbl>
#> 1 Mauna Loa CO2      1.24 0.00157
```



## 2 Exponential Smoothing (ETS) Forecasting Models

Forecasts produced using exponential smoothing methods are weighted averages of past observations, with the weights decaying exponentially as the observations get older.

The parameters are estimated by maximising the “likelihood”. The likelihood is the probability of the data arising from the specified model. AIC, AICc and BIC can be used here to determine which of the ETS models is most appropriate for a given time series (see output `glance(fit_ets)`).

The model selection is based on recognising key components of the time series (trend and seasonal) and the way in which these enter the smoothing method (e.g., in an additive, damped or multiplicative manner).

- Mauna Loa  $CO_2$  data best Models: ETS(M,A,A) & ETS(A,A,A)
- Basel Temperature data best Models: ETS(A,N,A), ETS(A,A,A), ETS(A,Ad,A) (close together). Best Forecast accuracy is with ETS(A,A,A), ETS(A,Ad,A).
- Basel Precipitation data best Models: ETS(A,N,A), ETS(A,Ad,A), ETS(A,A,A) (close together). Best Forecast accuracy is with ETS(A,A,A), ETS(A,Ad,A), ETS(A,N,A),

Trend term “N” for Basel Temperature/Precipitation corresponds to a “pure” exponential smoothing which results in a slope  $\beta = 0$ . This results in a forecast predicting a constant level. This does not fit to the result of the STL decomposition. Therefore best model choice is **ETS(A,A,A)**.

### Method Selection

*Error term:* either additive (“A”) or multiplicative (“M”).

Both methods provide identical point forecasts, but different prediction intervals and different likelihoods. AIC & BIC are able to select between the error types because they are based on likelihood.

Nevertheless, difference is for

- Mauna Loa  $CO_2$  not relevant and AIC/AICc/BIC values are only a little bit smaller for multiplicative errors. The prediction interval plots are fully overlapping.
- Basel Temperature AIC/AICc/BIC of additive error types are much better than the multiplicative ones.
- Basel Precipitation AIC/AICc/BIC of additive error types are much better than the multiplicative ones.

Note: For Basel Temperature and Precipitation Forecast plots the models ETS\_MAdA, ETS\_MMA, ETS\_MMA, ETS\_MNA are to be taken out since forecasts with multiplicative errors are exploding (forecast > 3 years impossible !!)

Therefore finally **Error term = “A”** is chosen in general.

*Trend term:* either none (“N”), additive (“A”), multiplicative (“M”) or damped variants (“Ad”, “Md”).

Note: Mauna Loa  $CO_2$  model ETS(A,Ad,A) fit plot shows to strong damping. For Basel Temperature model ETS(A,N,A) and ETS(A,Ad,A) are providing more or less the same forecast. This means that forecast remains on constant level since Trend “N” means “pure” exponential smoothing without trend (see above).

Therefore finally **Trend term = “A”** is chosen in general.

*Seasonal term:* either none (“N”), additive (“A”) or multiplicative (“M”).

For  $CO_2$  and Temperature Data we have a clear seasonal pattern and seasonal term adds always a (more or less) fix amount on level and trend component. Therefore “A” additive term is chosen. For Precipitation the seasonal pattern is only slight. Indeed, a multiplicative seasonal term results in “exploding” forecasts.

Since monthly data are strongly seasonal **seasonal term “A”** is chosen.

## 2.1 ETS Models and their componentes

ETS model with automatically selected  $ETS(A|M, N|A|M, N|A|M)$

```
#> [1] "model(ETS(count)) => provides default / best automatically chosen model"
#> # A mable: 1 x 3
#> # Key:      City, Measure [1]
#>   City      Measure      ETS
#>   <chr>     <fct>      <model>
#> 1 Mauna Loa CO2      <ETS(M,A,A)>
#> [1] "Mauna Loa CO2"
#> Series: count
#> Model: ETS(M,A,A)
#> Smoothing parameters:
#>   alpha = 0.5699886
#>   beta  = 0.006503521
#>   gamma = 0.08265732
#>
#> Initial states:
#>   l[0]    b[0]      s[0]    s[-1]    s[-2]    s[-3]    s[-4]
#> 320.4457 0.1017927 -0.8315672 -1.986214 -3.127605 -2.989192 -1.304845
#>   s[-5]    s[-6]    s[-7]    s[-8]    s[-9]    s[-10]    s[-11]
#> 0.6945233 2.128292 2.827021 2.446585 1.395346 0.7035553 0.04410057
#>
#> sigma^2: 0
#>
#>      AIC      AICc      BIC
#> 3102.618 3103.490 3180.465
#> # A tibble: 1 x 8
#>   City      Measure .model    AIC  AICc    BIC    MSE      MAE
#>   <chr>     <fct>   <chr>  <dbl> <dbl> <dbl> <dbl>   <dbl>
#> 1 Mauna Loa CO2     ETS    3103. 3103. 3180. 0.100 0.000674
```

Fit of different pre-defined  $ETS(A|M, N|A|M, N|A|M)$  models

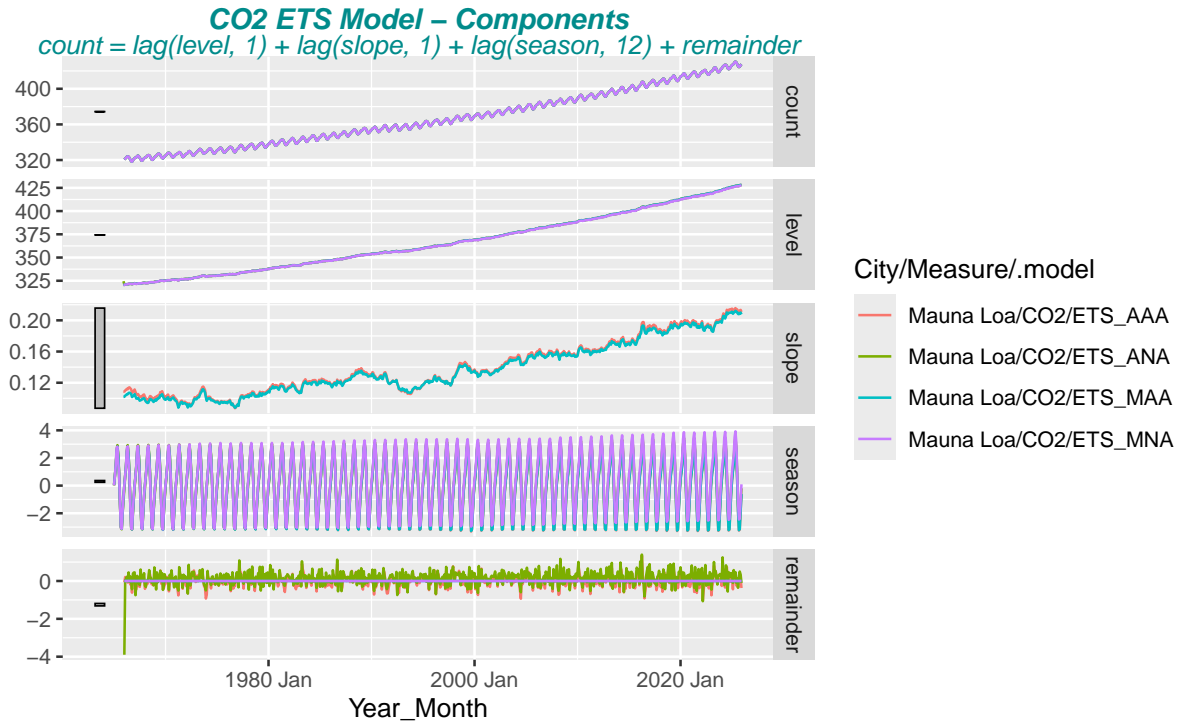
Model Selection by Information Criterion - lowest AIC, AICc, BIC

Best model fit with

- CV, AIC, AICc and BIC with the lowest values
- Adjusted  $R^2$  the model with the highest value.

```
#> # A tibble: 8 x 9
#>   City      Measure .model    AIC  AICc    BIC    MSE  AMSE      MAE
#>   <chr>     <fct>   <chr>  <dbl> <dbl> <dbl> <dbl> <dbl>   <dbl>
#> 1 Mauna Loa CO2     ETS_MAA 3103. 3103. 3180. 0.100 0.141 0.000674
#> 2 Mauna Loa CO2     ETS_AAA 3116. 3117. 3194. 0.100 0.140 0.248
#> 3 Mauna Loa CO2     ETS_MAdA 3161. 3162. 3244. 0.109 0.167 0.000703
#> 4 Mauna Loa CO2     ETS_AAdA 3167. 3168. 3250. 0.107 0.164 0.257
#> 5 Mauna Loa CO2     ETS_MNA 3362. 3363. 3431. 0.145 0.287 0.000823
#> 6 Mauna Loa CO2     ETS_MMA 3438. 3439. 3516. 0.153 0.187 0.000837
#> 7 Mauna Loa CO2     ETS_ANA 3452. 3452. 3520. 0.161 0.304 0.304
#> 8 Mauna Loa CO2     ETS_AMA 3538. 3539. 3616. 0.180 0.208 0.327
```





### 2.1.1 Residual Accuracy with one-step-ahead fitted residuals - check RMSE, MAE

Residual accuracy can be computed directly from models as the one-step-ahead fitted residuals are available. Select forecast models that minimises for lowest

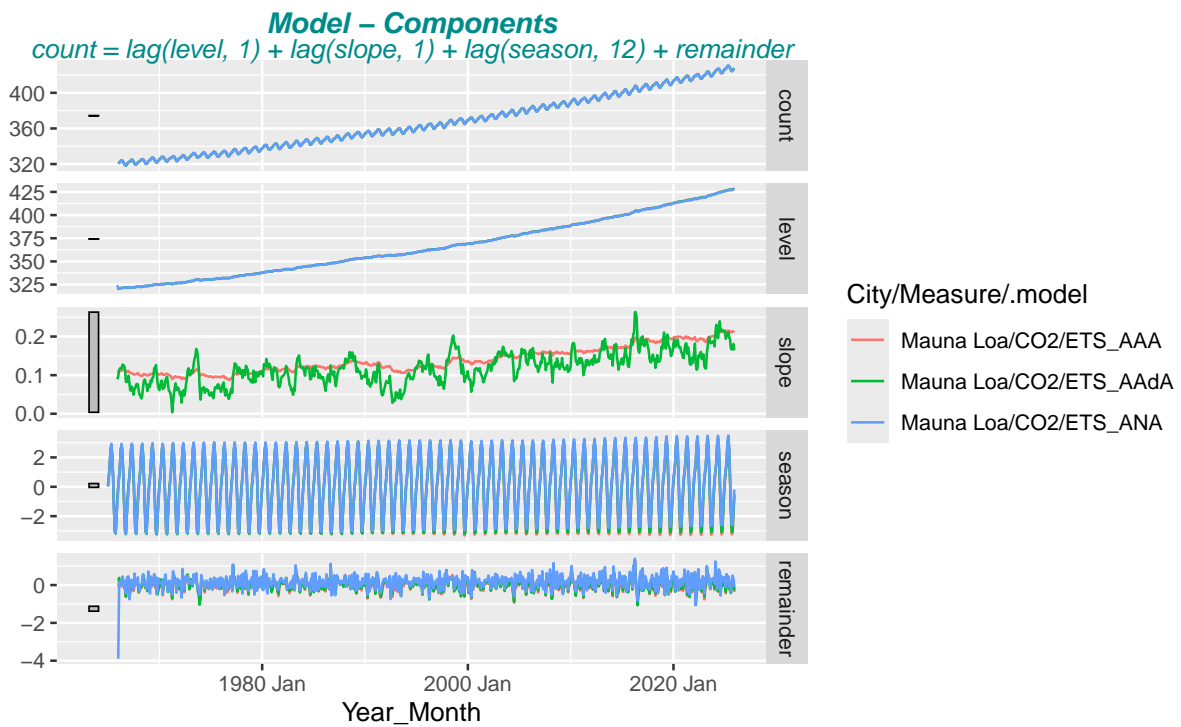
- MAE (Mean absolute error, will lead to forecasts of the median) and
- RMSE (Root mean squared error, lead to forecasts of the mean)

```
#> # A tibble: 8 x 7
#>   City      Measure .model  .type      ME  RMSE  MAE
#>   <chr>    <fct>   <chr>   <chr>    <dbl> <dbl> <dbl>
#> 1 Mauna Loa CO2     ETS_MAA Training 0.0226 0.317 0.247
#> 2 Mauna Loa CO2     ETS_AAA Training 0.0201 0.317 0.248
#> 3 Mauna Loa CO2     ETS_AAdA Training 0.0548 0.328 0.257
#> 4 Mauna Loa CO2     ETS_MAdA Training 0.0508 0.330 0.258
#> 5 Mauna Loa CO2     ETS_MNA Training 0.183  0.381 0.302
#> 6 Mauna Loa CO2     ETS_MMA Training 0.0229 0.391 0.304
#> 7 Mauna Loa CO2     ETS_ANA Training 0.151  0.401 0.304
#> 8 Mauna Loa CO2     ETS_AMA Training 0.0405 0.425 0.327
```

### 2.1.2 Ljung-Box Test - independence/white noise of the forecasts residuals

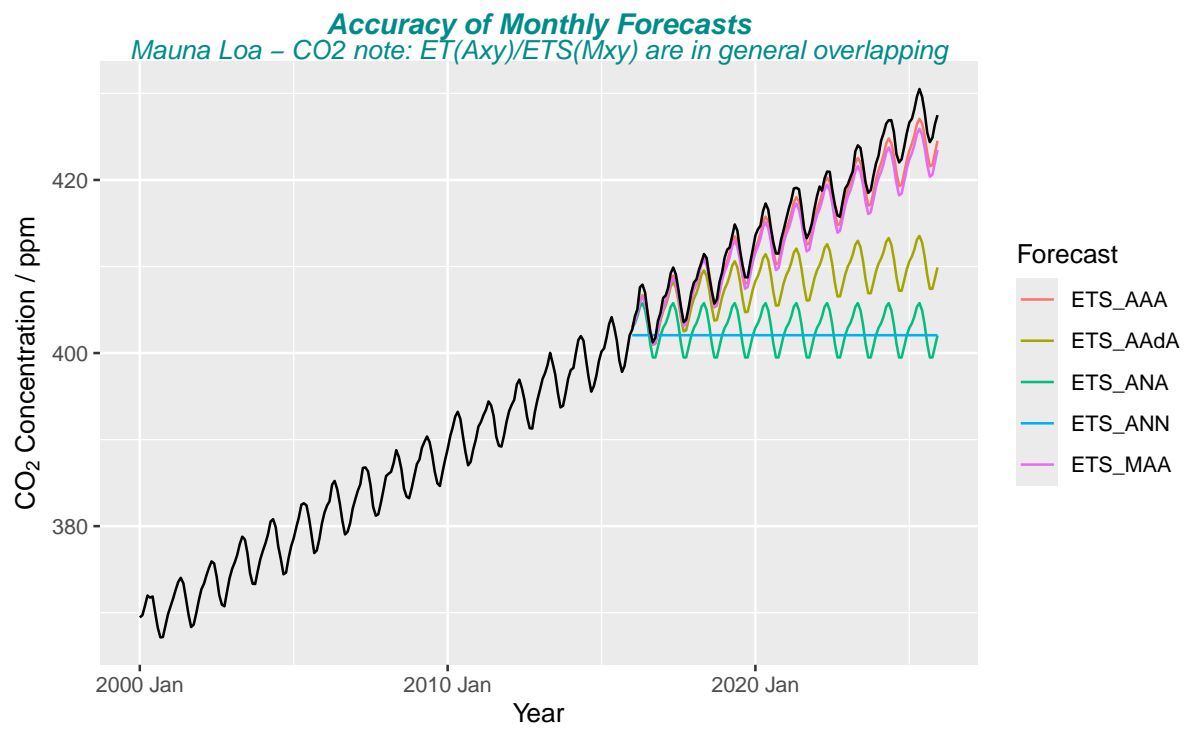
```
#> Null Hypothesis of independence/white noise for residuals - for p < 0.05: reject H_0
#> # A tibble: 8 x 5
#>   City      Measure .model  lb_stat lb_pvalue
#>   <chr>    <fct>   <chr>    <dbl>    <dbl>
#> 1 Mauna Loa CO2     ETS_MAA    35.9  0.0226
#> 2 Mauna Loa CO2     ETS_AAA    37.0  0.0167
#> 3 Mauna Loa CO2     ETS_MNA    44.7  0.00191
#> 4 Mauna Loa CO2     ETS_ANA    47.8  0.000735
#> 5 Mauna Loa CO2     ETS_AAdA    49.4  0.000438
#> 6 Mauna Loa CO2     ETS_MAdA    52.1  0.000182
#> 7 Mauna Loa CO2     ETS_AMA   324.    0
#> 8 Mauna Loa CO2     ETS_MMA   292.    0
```

### 2.1.3 ETS Models - components of ETS(A,N,A), ETS(A,A,A), ETS(A,Ad,A), models



### 2.1.4 Forecast Accuracy with Training/Test Data

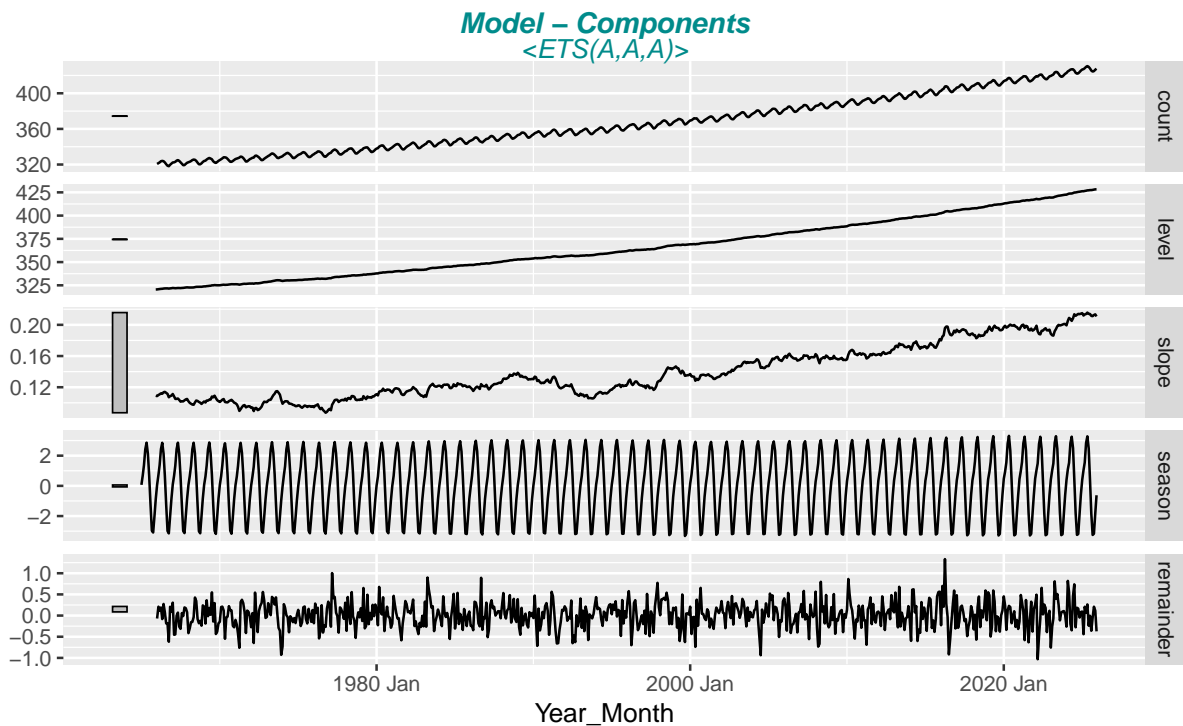
```
#> # A tibble: 5 x 7
#>   .model City      Measure .type    ME  RMSE  MAE
#>   <chr>   <chr>      <fct>   <chr> <dbl> <dbl> <dbl>
#> 1 ETS_AAA Mauna Loa CO2      Test  1.23  1.54  1.25
#> 2 ETS_MAA Mauna Loa CO2      Test  1.93  2.26  1.93
#> 3 ETS_AAdA Mauna Loa CO2      Test  7.21  8.94  7.21
#> 4 ETS_ANA Mauna Loa CO2      Test 12.6  14.6  12.6
#> 5 ETS_ANN Mauna Loa CO2      Test 13.3  15.3  13.3
```

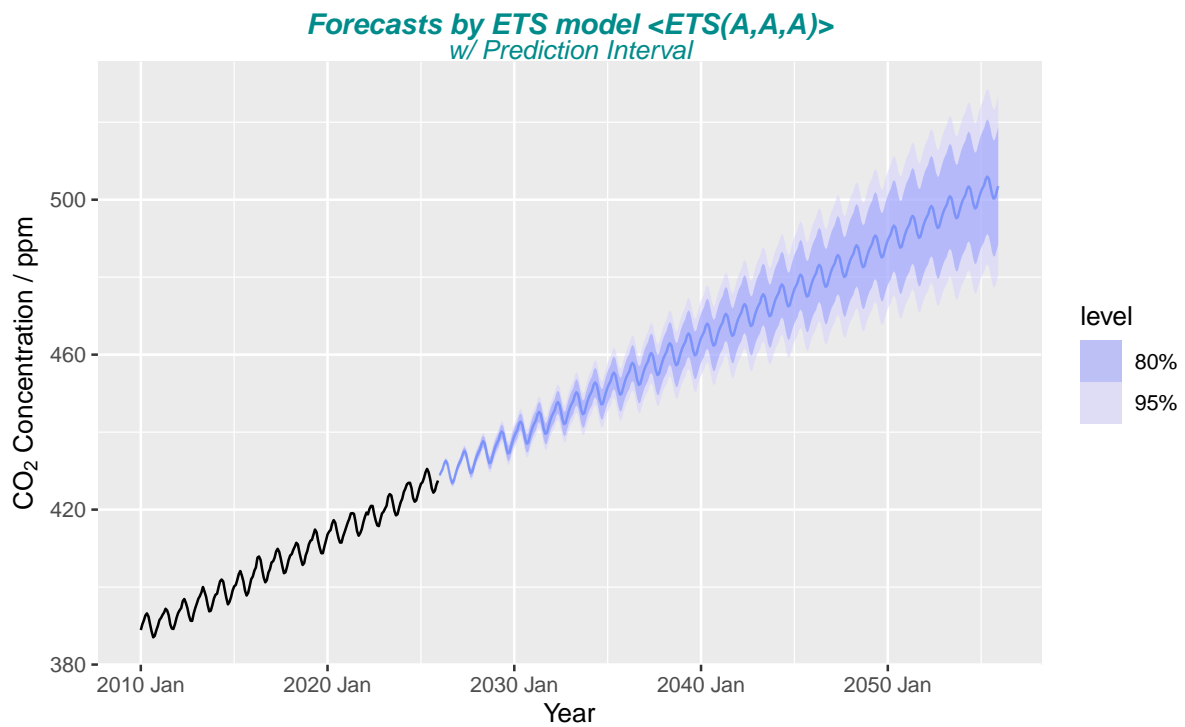


## 2.2 Forecasting with selected ETS model <ETS(A,A,A)>

### 2.2.1 Forecast Plot of selected ETS model

```
#> Provide model coefficients by report(fit_model)
#> Series: count
#> Model: ETS(A,A,A)
#> Smoothing parameters:
#>   alpha = 0.5343921
#>   beta  = 0.007020123
#>   gamma = 0.0926534
#>
#> Initial states:
#>   l[0]    b[0]      s[0]    s[-1]    s[-2]    s[-3]    s[-4]    s[-5]
#> 320.5115 0.1090902 -0.8747296 -2.013688 -3.089998 -2.969892 -1.23776 0.6645857
#>   s[-6]    s[-7]    s[-8]    s[-9]    s[-10]    s[-11]
#> 2.107803 2.862666 2.422816 1.362172 0.6973195 0.06870423
#>
#> sigma^2: 0.1026
#>
#>      AIC      AICc      BIC
#> 3115.822 3116.694 3193.669
```

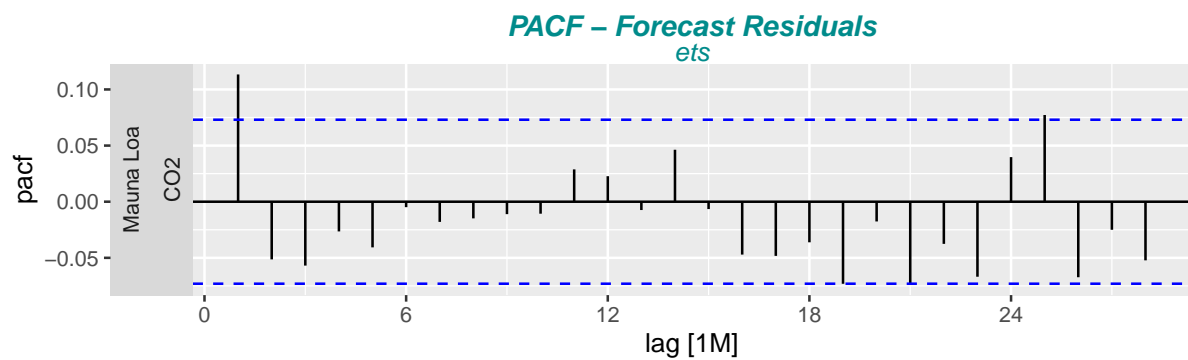
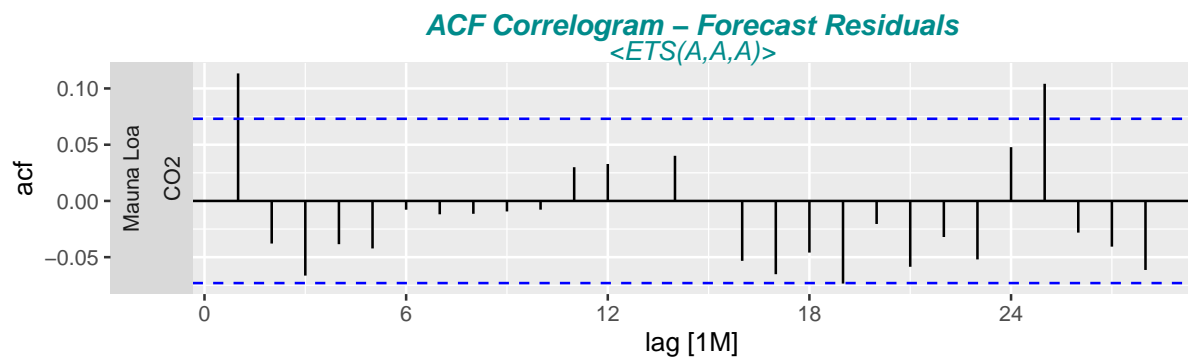


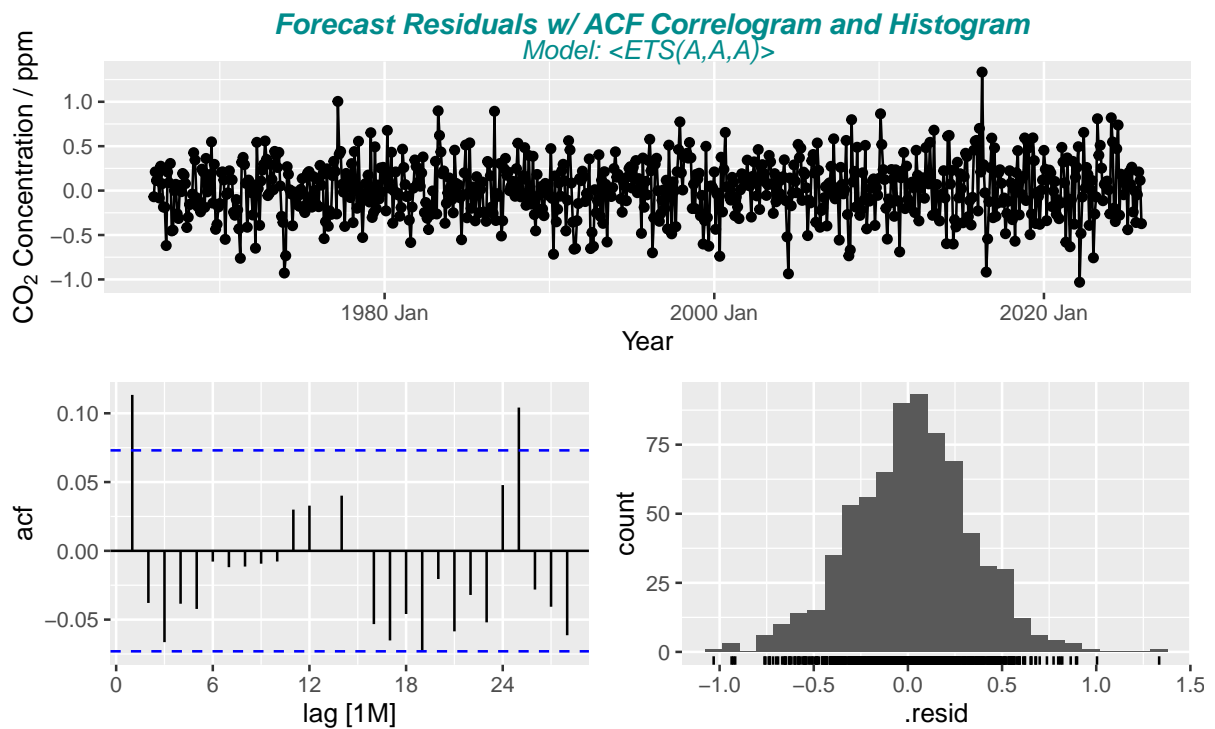


## 2.2.2 Residual Stationarity

Required checks to be ready for forecasting:

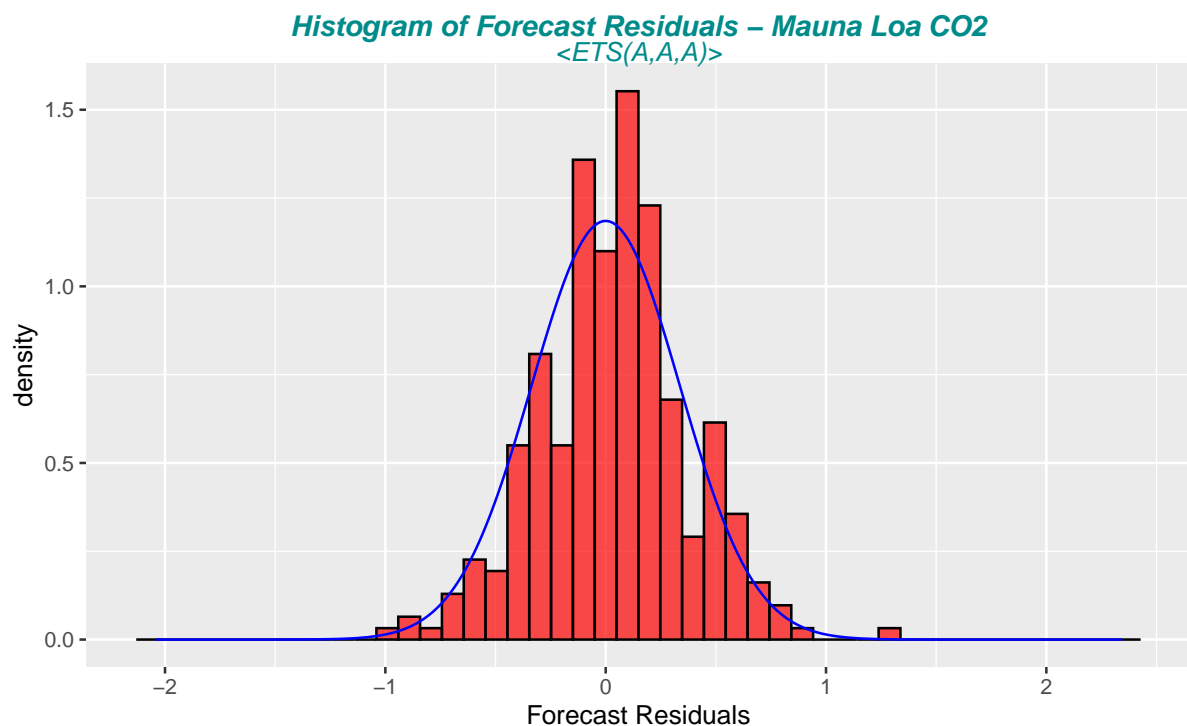
- ACF Forecast Residual: all spikes are within the significance limits, so the residuals appear to be white noise
- The Ljung-Box test also shows that the residuals have no remaining autocorrelations
- Forecast Residuals are more or less normally distributed with roughly centred on zero





### 2.2.3 Ljung-Box Test and Histogram of forecast residuals with overlaid normal curve

```
#> Null Hypothesis of independence/white noise for residuals - for p < 0.05: reject H_0
#> # A tibble: 1 x 5
#>   City      Measure .model lb_stat lb_pvalue
#>   <chr>    <fct>    <chr>   <dbl>   <dbl>
#> 1 Mauna Loa CO2     ets      27.2    0.295
```



### 3 ARIMA Forecasting Models - AutoRegressive-Integrated Moving Average

Exponential smoothing and ARIMA (AutoRegressive-Integrated Moving Average) models are the two most widely used approaches to time series forecasting, and provide complementary approaches to the problem.

While exponential smoothing models are based on a description of the trend and seasonality in the data, ARIMA models aim to describe the autocorrelations in the data.

#### 3.1 Seasonal ARIMA models

Non-seasonal ARIMA models are generally denoted  $ARIMA(p,d,q)$  where parameters  $p$ ,  $d$ , and  $q$  are non-negative integers, \*  $p$  is the order (number of time lags) of the autoregressive model \*  $d$  is the degree of differencing (number of times the data have had past values subtracted) \*  $q$  is the order of the moving-average model of past forecast errors .

The value of  $d$  has an effect on the prediction intervals — the higher the value of  $d$ , the more rapidly the prediction intervals increase in size. For  $d=0$ , the point forecasts are equal to the mean of the data and the long-term forecast standard deviation will go to the standard deviation of the historical data, so the prediction intervals will all be essentially the same.

Seasonal ARIMA models are usually denoted  $ARIMA(p,d,q)(P,D,Q)m$ , where  $m$  refers to the number of periods in each season, and the uppercase  $P,D,Q$  refer to the autoregressive, differencing, and moving average terms for the seasonal part of the ARIMA model.

*ARIMA(pdq)(PDQ) model with automatically selected (pdq)(PDQ) values*

**Fit of different pre-defined *ARIMA(pdq)(PDQ)* models**

```
#> Model Selection by Information Criterion - lowest AIC, AICc, BIC
#> choose p, q parameter accordingly - but only for same d, D values
#> # A tibble: 8 x 8
#>   City      Measure .model      sigma2 log_lik   AIC   AICc   BIC
#>   <chr>    <fct>    <chr>      <dbl>   <dbl> <dbl> <dbl> <dbl>
#> 1 Mauna Loa CO2    arima_111_011 0.101   -192.  393.  393.  411.
#> 2 Mauna Loa CO2    arima_012_011 0.101   -193.  393.  394.  412.
#> 3 Mauna Loa CO2    arima_111_012 0.101   -192.  395.  395.  418.
#> 4 Mauna Loa CO2    arima_211_011 0.101   -192.  395.  395.  418.
#> 5 Mauna Loa CO2    arima_012_112 0.101   -193.  397.  398.  425.
#> 6 Mauna Loa CO2    arima_x1y_m1n 0.101   -193.  397.  398.  425.
#> 7 Mauna Loa CO2    arima_210_110 0.133   -285.  579.  579.  597.
#> 8 Mauna Loa CO2    arima_100_110_c 0.133   -286.  582.  583.  605.
```

Good models are obtained by minimising the AIC, AICc or BIC (see `glance(fit_arima)` output). The preference is to use the AICc to select  $p$  and  $q$ .

These information criteria tend not to be good guides to selecting the appropriate order of differencing ( $d$ ) of a model, but only for selecting the values of  $p$  and  $q$ . This is because the differencing changes the data on which the likelihood is computed, making the AIC values between models with different orders of differencing not comparable.

##### 3.1.1 Residual Accuracy with one-step-ahead fitted residuals - check RMSE, MAE

Residual accuracy can be computed directly from models as the one-step-ahead fitted residuals are available. Select forecast models that minimises for lowest

- MAE (Mean absolute error, will lead to forecasts of the median) and
- RMSE (Root mean squared error, lead to forecasts of the mean)

```
#> # A tibble: 8 x 7
#>   City      Measure .model      .type      ME  RMSE  MAE
#>   <chr>    <fct>    <chr>      <chr>    <dbl> <dbl> <dbl>
```

```
#> 1 Mauna Loa CO2      arima_111_012 Training 0.0262 0.314 0.243
#> 2 Mauna Loa CO2      arima_211_011 Training 0.0263 0.314 0.243
#> 3 Mauna Loa CO2      arima_111_011 Training 0.0261 0.314 0.243
#> 4 Mauna Loa CO2      arima_012_112 Training 0.0253 0.314 0.243
#> 5 Mauna Loa CO2      arima_x1y_m1n Training 0.0253 0.314 0.243
#> 6 Mauna Loa CO2      arima_012_011 Training 0.0253 0.314 0.243
#> 7 Mauna Loa CO2      arima_100_110_c Training 0.00664 0.361 0.287
#> 8 Mauna Loa CO2      arima_200_011 Training 0.00664 0.361 0.287
```

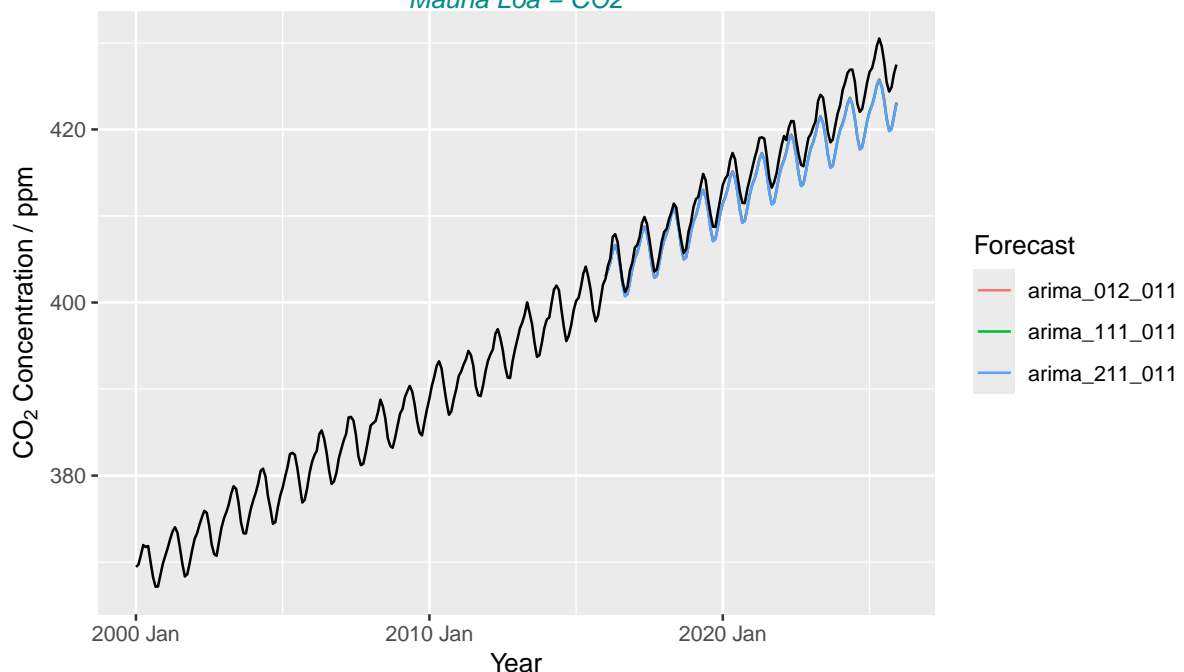
### 3.1.2 Ljung-Box Test - independence/white noise of the forecasts residuals

```
#> Null Hypothesis of independence/white noise for residuals - for p < 0.05: reject H_0
#> # A tibble: 8 x 5
#>   City      Measure .model      lb_stat lb_pvalue
#>   <chr>      <fct>   <chr>      <dbl>   <dbl>
#> 1 Mauna Loa CO2      arima_211_011      16.5 7.42e- 1
#> 2 Mauna Loa CO2      arima_111_011      16.5 7.42e- 1
#> 3 Mauna Loa CO2      arima_111_012      16.6 7.36e- 1
#> 4 Mauna Loa CO2      arima_012_011      16.7 7.31e- 1
#> 5 Mauna Loa CO2      arima_012_112      16.7 7.31e- 1
#> 6 Mauna Loa CO2      arima_x1y_m1n      16.7 7.31e- 1
#> 7 Mauna Loa CO2      arima_100_210      79.5 9.96e- 9
#> 8 Mauna Loa CO2      arima_100_110_c     105. 4.08e-13
```

### 3.1.3 Forecast Accuracy with Training/Test Data

```
#> # A tibble: 3 x 7
#>   .model      City      Measure .type      ME  RMSE  MAE
#>   <chr>      <chr>      <fct>   <chr> <dbl> <dbl> <dbl>
#> 1 arima_012_011 Mauna Loa CO2      Test    2.07  2.42  2.07
#> 2 arima_111_011 Mauna Loa CO2      Test    2.12  2.47  2.13
#> 3 arima_211_011 Mauna Loa CO2      Test    2.18  2.53  2.18
```

**Accuracy of Monthly Forecasts w/ Training and Test data**  
Mauna Loa – CO2

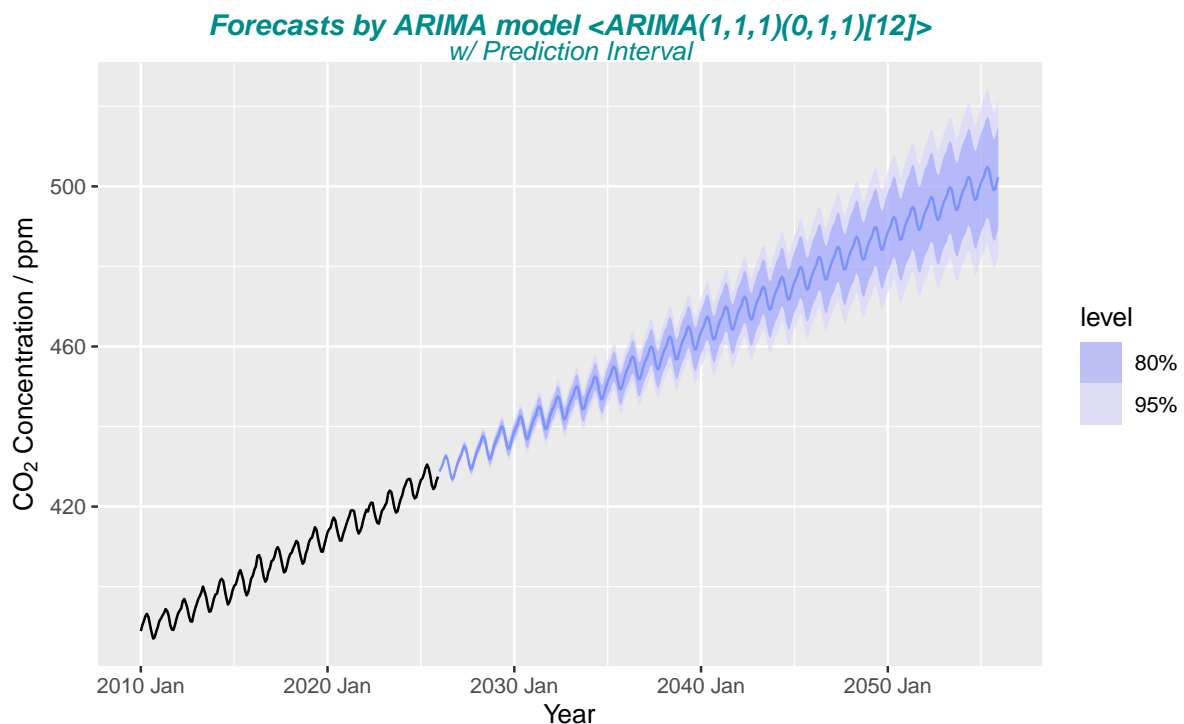




## 3.2 CO2 - Forecasting with selected ARIMA model <ARIMA(1,1,1)(0,1,1)[12]>

### 3.2.1 Forecast Plot of selected ARIMA model

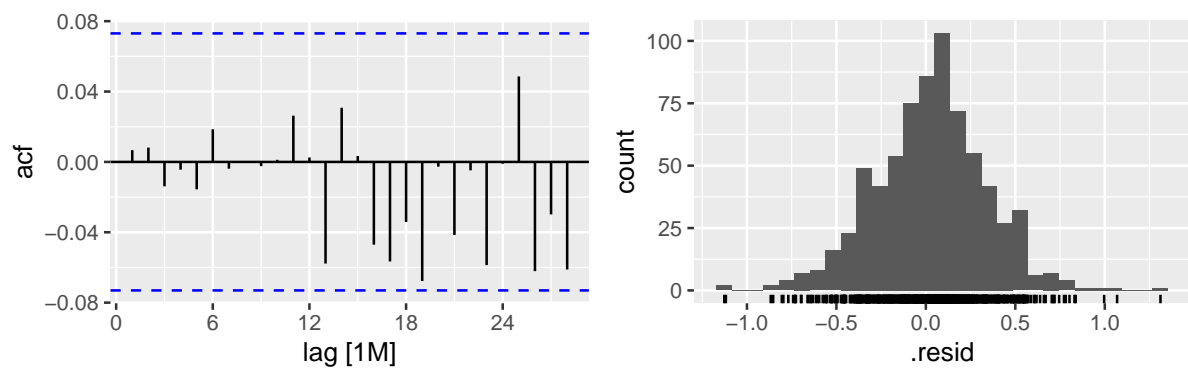
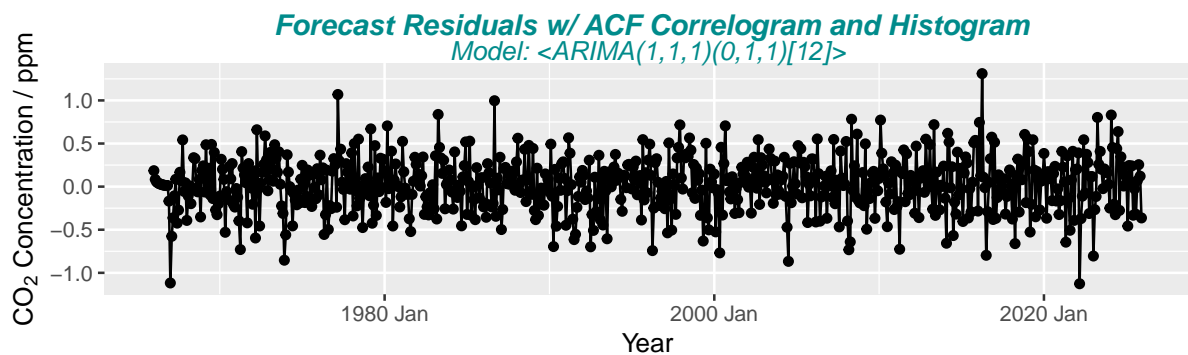
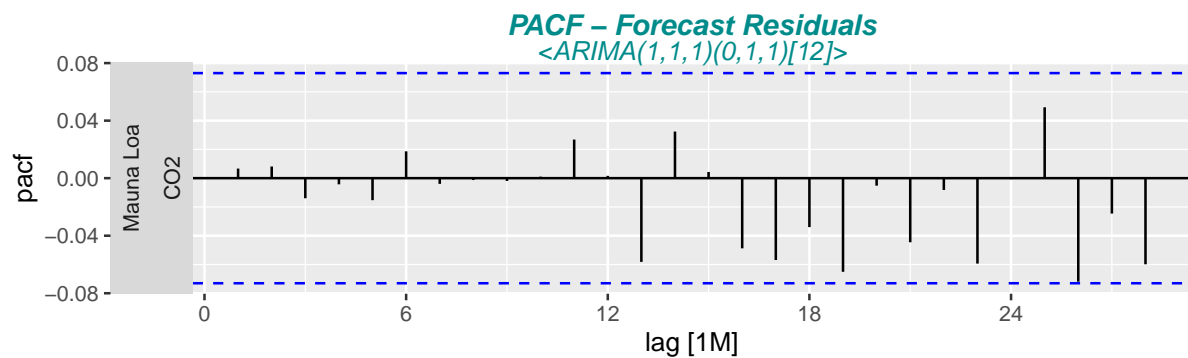
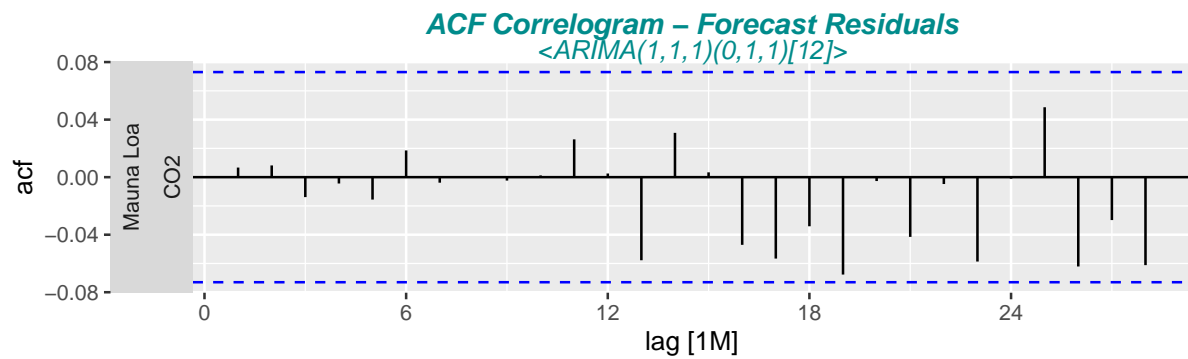
```
#> Provide model coefficients by report(fit_model)
#> Series: count
#> Model: ARIMA(1,1,1)(0,1,1)[12]
#>
#> Coefficients:
#>          ar1          ma1          sma1
#>         0.2250      -0.5716      -0.8758
#> s.e.  0.0901      0.0754      0.0189
#>
#> sigma^2 estimated as 0.1007:  log likelihood=-192.48
#> AIC=392.96   AICc=393.01   BIC=411.2
```



### 3.2.2 Residual Stationarity

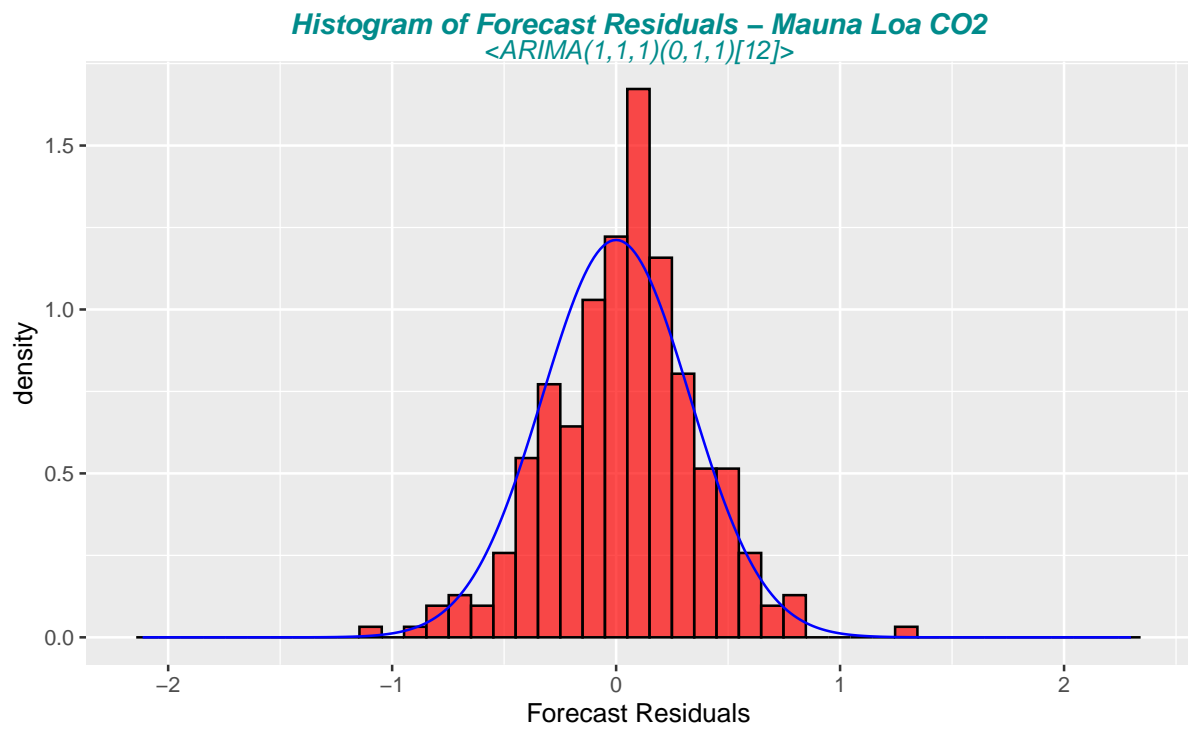
Required checks to be ready for forecasting:

- ACF Forecast Residual: all spikes are within the significance limits, so the residuals appear to be white noise
- The Ljung-Box test also shows that the residuals have no remaining autocorrelations
- Forecast Residuals are more or less normally distributed with roughly centred on zero



### 3.2.3 Ljung-Box Test and Histogram of forecast residuals with overlaid normal curve

```
#> Null Hypothesis of independence/white noise for residuals - for p < 0.05: reject H_0
#> # A tibble: 1 x 5
#>   City      Measure .model lb_stat lb_pvalue
#>   <chr>    <fct>    <chr>   <dbl>   <dbl>
#> 1 Mauna Loa CO2      arima    19.9    0.528
```



## 4 ARIMA vs ETS

In particular, all ETS models are non-stationary, while some ARIMA models are stationary.

The ETS models with seasonality or non-damped trend or both have two unit roots (i.e., they need two levels of differencing to make them stationary). All other ETS models have one unit root (they need one level of differencing to make them stationary).

We compare for the chosen ETS resp. ARIMA model the RMSE / MAE values. Lower values indicate a more accurate model based on the test set RMSE, ..., MASE.

- Residual Accuracy with one-step-ahead fitted residuals
- Forecast Accuracy with Training/Test Data

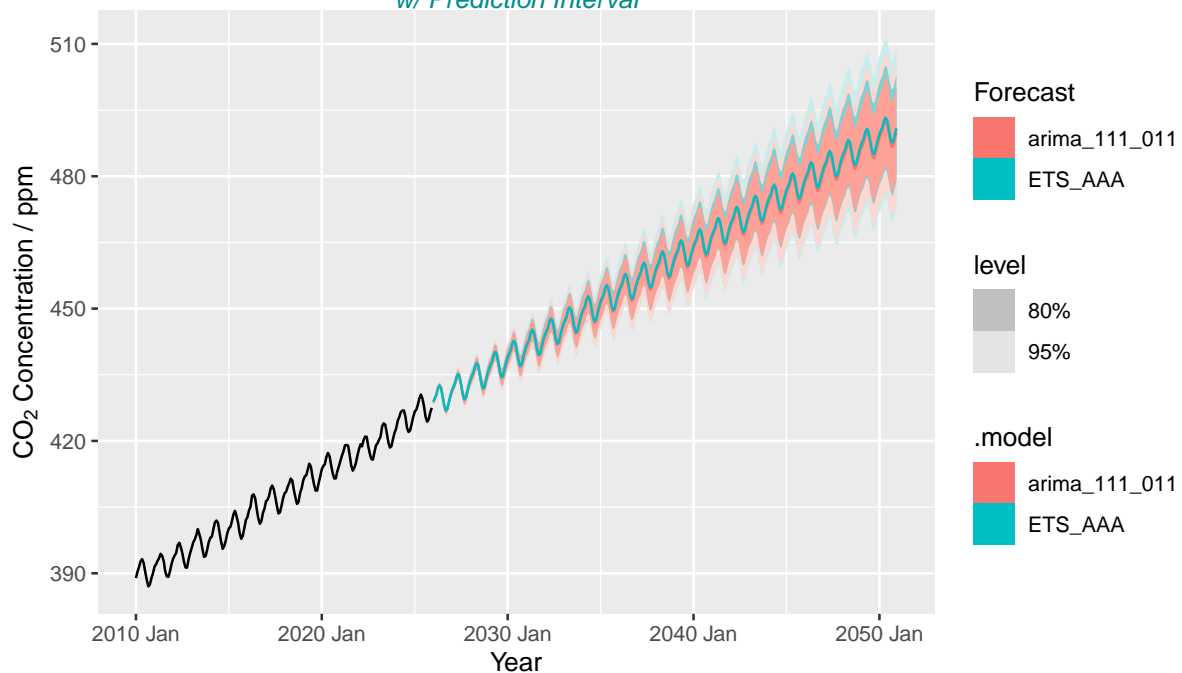
Note: a good fit to training data is never an indication that the model will forecast well. Therefore the values of the Forecast Accuracy are the more relevant one.

### 4.0.1 Comparing Residual and Forecast Accuracy of selected ETS and ARIMA model

```
#> # A tibble: 4 x 9
#>   City      Measure .model      .type      RMSE      MAE      MAPE      MASE      RMSSE
#>   <chr>    <fct>    <chr>      <chr>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>
#> 1 Mauna Loa CO2      arima    Training 0.314    0.243    0.0664    0.135    0.160
#> 2 Mauna Loa CO2      ets      Training 0.317    0.248    0.0676    0.138    0.161
#> 3 Mauna Loa CO2      ETS_AAA    Test     1.54     1.25     0.298     0.769    0.872
#> 4 Mauna Loa CO2      arima_111_011 Test     2.47     2.13     0.507     1.31     1.40
```

### 4.0.2 Forecast Plot of selected ETS and ARIMA model

**Forecasts by ETS <ETS(A,A,A)> and ARIMA model <ARIMA(1,1,1)(0,1,1)[12]>**  
w/ Prediction Interval



Forecasts by ETS <ETS(A,A,A)> and ARIMA model <ARIMA(1,1,1)(0,1,1)[12]>  
w/ Prediction Interval

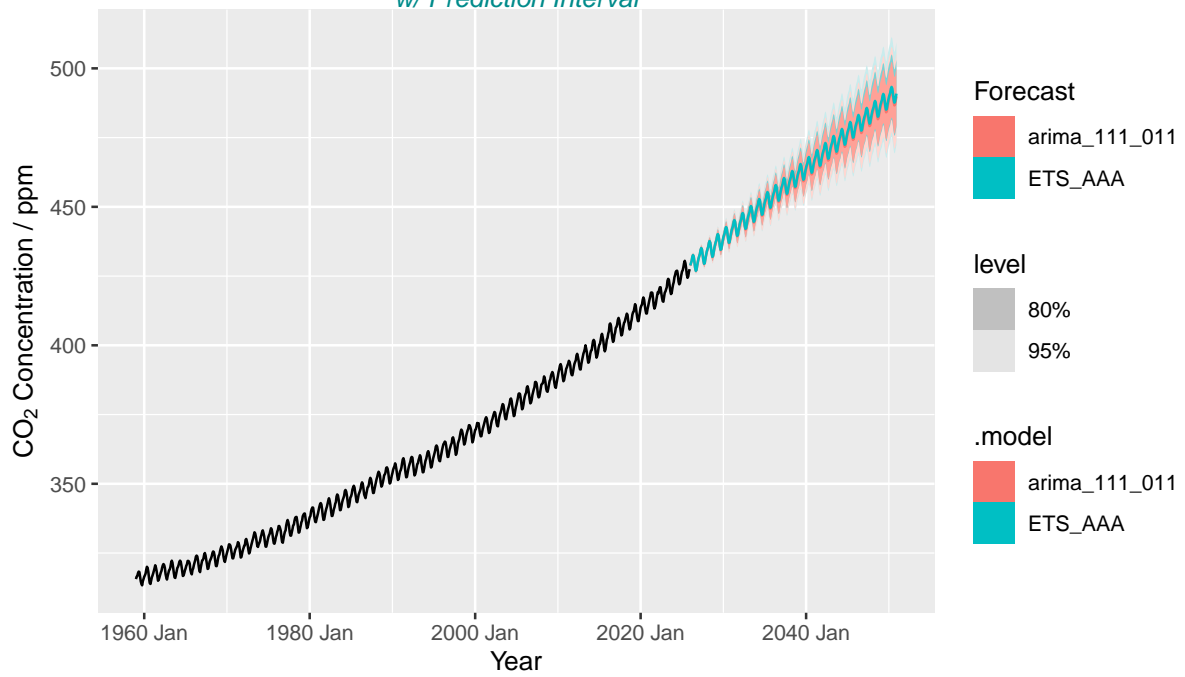


Table 1: Mean values for the given time periods; Units: Temperature (degree C), Precipitation (mm/Month), CO2 (ppm)

Period_Time	CO2
1931-1960	316.4
1961-1990	333.5
1991-2020	381.9
2021-2025	421.6

Table 2: Mean Yearly ARIMA and ETS Forecast values (next 25 years); Units: Temperature (degree C), Precipitation (mm/Month), CO2 (ppm)

City	Measure	Year	ETS_AAA	arima_111_011
Mauna Loa	CO2	2026	429.73	429.73
Mauna Loa	CO2	2030	439.85	439.70
Mauna Loa	CO2	2035	452.50	452.15
Mauna Loa	CO2	2040	465.15	464.60
Mauna Loa	CO2	2045	477.80	477.05
Mauna Loa	CO2	2050	490.45	489.50

Table 3: Forecast increase/decrease over the next 25 years; Units: Temperature (degree C), Precipitation (mm/Month), CO2 (ppm)

Measure	Year.x	Year.y	ETS.x	ARIMA.x	ETS.y	ARIMA.y	Delta_ETs	Delta_ARIMA
CO2	2026	2050	429.73	429.73	490.45	489.5	60.72	59.77

Table 4: Forecast increase/decrease over the next 25 years; Units: Temperature (degree C), Precipitation (mm/Month), CO2 (ppm)

Measure	Month	Year.x	Year.y	ETS.x	ARIMA.x	ETS.y	ARIMA.y	Delta_ETs	Delta_ARIMA
CO2	Jan	2026	2050	428.86	428.79	489.58	488.58	60.72	59.79
CO2	Feb	2026	2050	429.73	429.70	490.45	489.48	60.72	59.77
CO2	Mrz	2026	2050	430.43	430.43	491.15	490.19	60.72	59.77
CO2	Apr	2026	2050	431.93	431.93	492.65	491.69	60.72	59.77
CO2	Mai	2026	2050	432.59	432.61	493.31	492.38	60.72	59.76
CO2	Jun	2026	2050	432.00	432.08	492.72	491.84	60.72	59.76
CO2	Jul	2026	2050	430.17	430.28	490.89	490.04	60.72	59.76
CO2	Aug	2026	2050	428.10	428.14	488.82	487.91	60.72	59.76
CO2	Sep	2026	2050	426.90	426.87	487.62	486.63	60.72	59.76
CO2	Okt	2026	2050	427.19	427.14	487.91	486.91	60.72	59.76
CO2	Nov	2026	2050	428.75	428.70	489.47	488.47	60.72	59.76
CO2	Dez	2026	2050	430.16	430.11	490.88	489.87	60.72	59.76

## 5 Yearly Data Forecasts with ARIMA and ETS

For yearly data the seasonal monthly data are replaced by the yearly average data. Therefore the seasonal component of the ETS and ARIMA model are to be taken out.

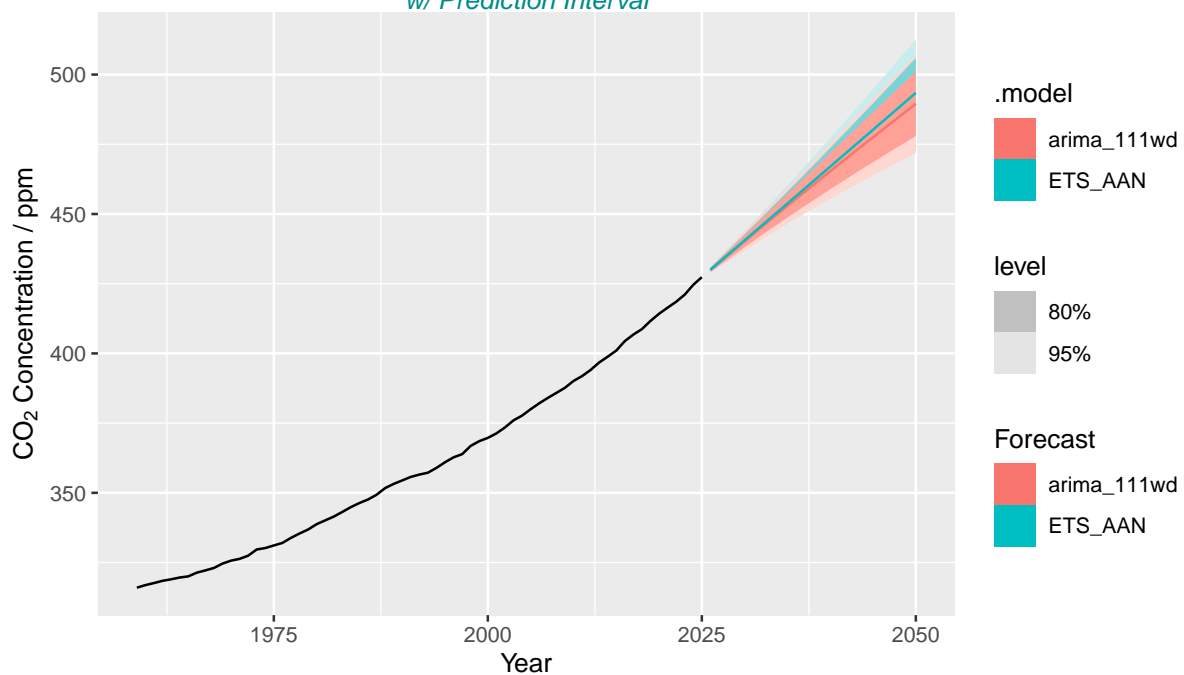
The ETS model  $\langle ETS(A, A, N) \rangle$  with seasonal term change “A”  $\rightarrow$  “N” is chosen. For ARIMA models the seasonal term (P,D,Q)<sub>m</sub> has to be taken out and an optimal ARIMA(p,1,q) with one differencing (d=1) is selected. However, for Mauna Loa two times differencing had to be selected  $\$CO\_2 \langle ARIMA(0,2,1) \text{ w/ poly} \rangle$ . For Temperature and Precipitation the same model as for monthly data can be taken by leaving out the seasonal term  $\langle ARIMA(0,1,2) \text{ w/ drift} \rangle$ .

### 5.0.1 Comparing Residual and Forecast Accuracy of selected ETS and ARIMA model

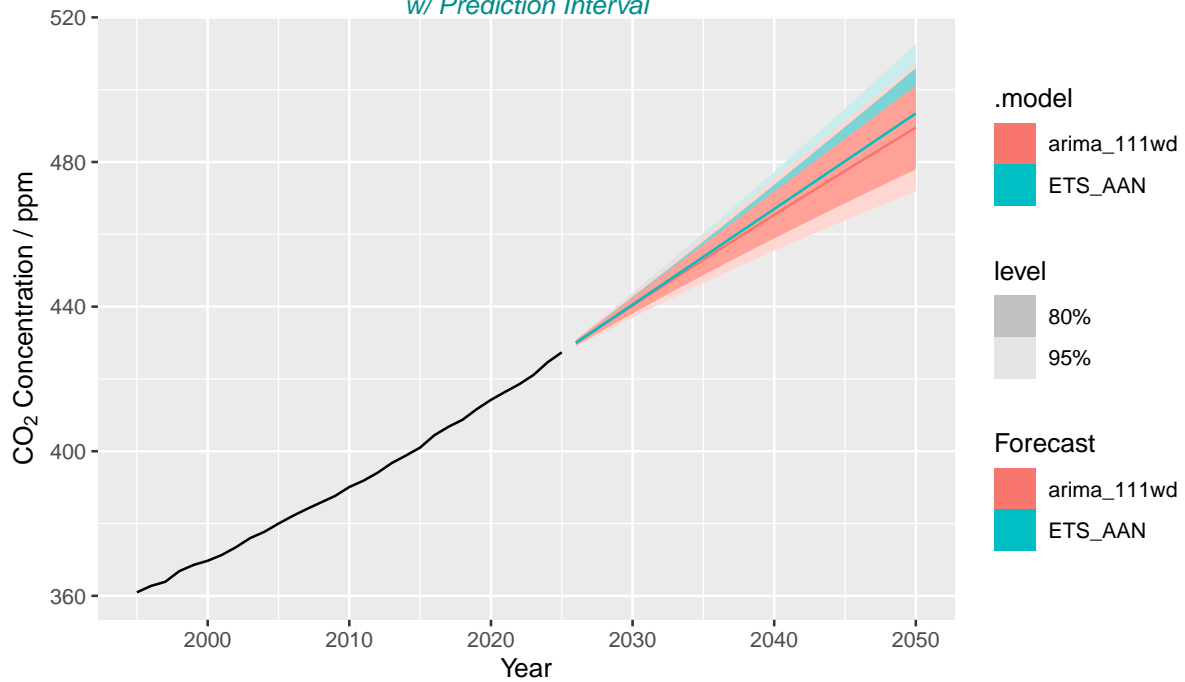
### 5.0.2 Forecast Plot of selected ETS and ARIMA model

```
#> selected: ETS_AAN and arima_111 w/ drift
```

Yearly Forecasts by ETS  $\langle ETS(A,A,N) \rangle$  and ARIMA model  $\langle ARIMA(1,1,1) \text{ w/ drift} \rangle$   
w/ Prediction Interval



## Early Forecasts by ETS <ETS(A,A,N)> and ARIMA model <ARIMA(1,1,1) w/ drift> w/ Prediction Interval



```
#> # A tibble: 2 x 13
#>   City      Measure .model sigma2 log_lik  AIC  AICc  BIC  MSE  AMSE  MAE
#>   <chr>    <fct>   <chr>  <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
#> 1 Mauna Loa CO2    arima~ 0.282 -45.4  98.9  99.6 107.  NA    NA    NA
#> 2 Mauna Loa CO2    ETS_A~ 0.277 -82.2 174.  176. 185. 0.258 0.619 0.385
#> # i 2 more variables: ar_roots <list>, ma_roots <list>
```

Ljung-Box Test - independence/white noise of the forecasts residuals

```
#> # A tibble: 1 x 3
#>   .model lb_stat lb_pvalue
#>   <chr>   <dbl>   <dbl>
#> 1 ETS_AAN 66.4 0.00000756
#> # A tibble: 1 x 3
#>   .model lb_stat lb_pvalue
#>   <chr>   <dbl>   <dbl>
#> 1 arima_111wd 65.6 0.00000172
```



Table 5: Mean Yearly ARIMA and ETS Forecast values of mean yearly data (next 25 years); Units: Temperature (degree C), Precipitation (mm/Month), CO2 (ppm)

City	Measure	Year	ETS_AAN	arima_111wd
Mauna Loa	CO2	2026	429.99	429.95
Mauna Loa	CO2	2030	440.57	440.27
Mauna Loa	CO2	2035	453.79	452.92
Mauna Loa	CO2	2040	467.00	465.34
Mauna Loa	CO2	2045	480.22	477.53
Mauna Loa	CO2	2050	493.44	489.52

Table 6: Forecast increase/decrease over the next 25 years; Units: Temperature (degree C), Precipitation (mm/Month), CO2 (ppm)

Measure	Year.x	Year.y	ETS.x	ARIMA.x	ETS.y	ARIMA.y	Delta_ETs	Delta_ARIMA
CO2	2026	2050	429.99	429.95	493.44	489.52	63.45	59.56