Climate Data Forecasting -

Atmospheric CO_2 Concentration / Temperature / Precipitation

Wolfgang Vollmer

2021-04-05

Contents

1 Forecasting of Mauna Loa - Atmospheric Carbon Dioxide Analysis 2

1.1 Stationarity and differencing . . . 2

1.1.1 Ljung-Box Test - independence/white noise of the time series . . . 3

1.1.2 Unitroot KPSS Test - fix number of seasonal differences/differences required . . . 3

1.1.3 ACF Plots of Differences . . . 4

1.1.4 Time Series, ACF and PACF Plots of Differences - for ARIMA p, q check . . . 5

2 ExponenTial Smoothing (ETS) Forecasting Models 6

2.1 ETS Models and their componentes . . . 7

2.1.1 Residual Accuracy with one-step-ahead fitted residuals - check RMSE, MAE . . . 8

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 . 9

2.1.4 Forecast Accuracy with Training/Test Data . . . 9

2.2 Forecasting with selected ETS model <ETS(A,A,A)> . . . 10

2.2.1 Forecast Plot of selected ETS model . . . 10

2.2.2 Residual Stationarity . . . 11

2.2.3 Histogram of forecast residuals with overlaid normal curve . . . 12

3 ARIMA Forecasting Models - AutoRegressive-Integrated Moving Average 13

3.1 Seasonal ARIMA models . . . 13

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 . . . 15

3.1.3 Forecast Accuracy with Training/Test Data . . . 15

3.2 CO2 - Forecasting with selected ARIMA model <ARIMA(1,1,1)(0,1,2)[12]> . . . 16

3.2.1 Forecast Plot of selected ARIMA model . . . 16

3.2.2 Residual Stationarity . . . 17

3.2.3 Histogram of forecast residuals with overlaid normal curve . . . 18

4 ARIMA vs ETS **19**
4.0.1 Comparing Residual and Forecast Accuracy of selected ETS and ARIMA model . 19
4.0.2 Forecast Plot of selected ETS and ARIMA model 19
4.0.3 Ljung-Box Test - independence/white noise of the forecasts residuals 21

5 Yearly Data Forecasts with ARIMA and ETS **21**
5.0.1 Comparing Residual and Forecast Accuracy of selected ETS and ARIMA model . 21
5.0.2 Forecast Plot of selected ETS and ARIMA model 22
5.0.3 Ljung-Box Test - independence/white noise of the forecasts residuals 23

6 Backup **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 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

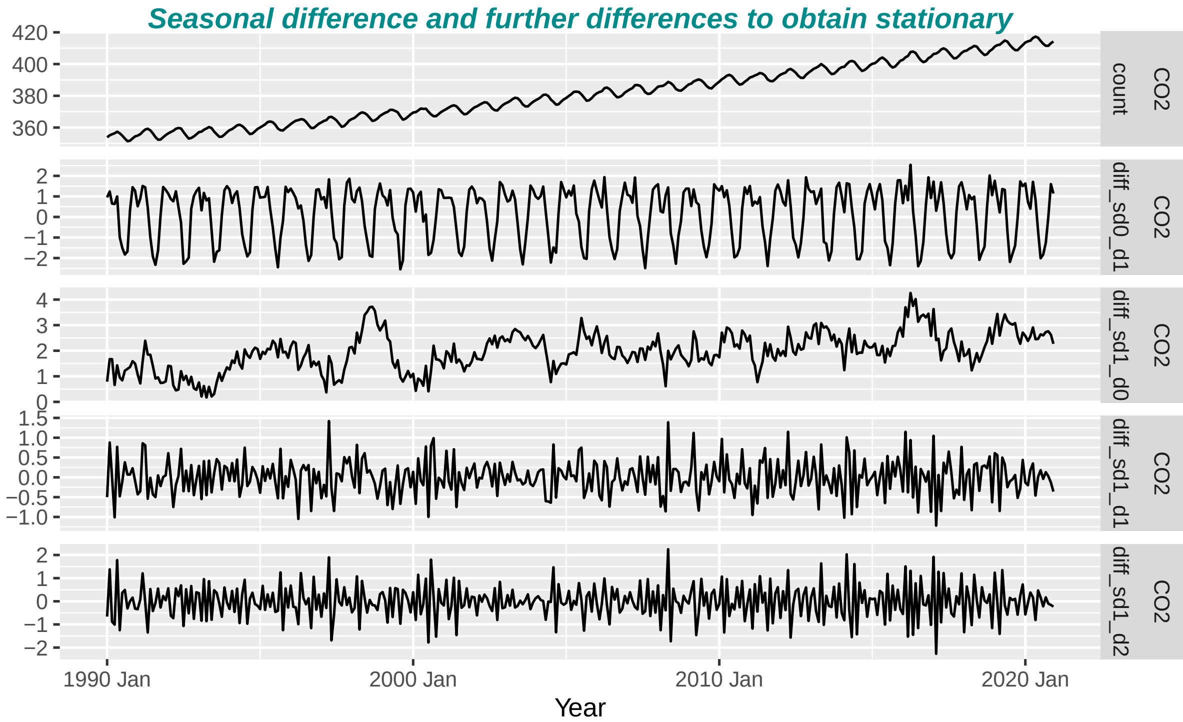
The model fit residuals have to be stationary. For good forecasting this has to be verified with residual diagnostics.

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 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 rsp. ARIMA models. There we have to check whether the forecast errors are normally distributed with mean zero

Null Hypothesis of independence/white noise in a given time series
=> H_0 to be rejected for $p < \alpha = 0.05$
=> data in the given time series are dependent
=> even differenced data are dependent if $p < \alpha = 0.05$
=> independence/white noise of residuals of fitted models to be verified

```
#> Ljung-Box test with (count), w/o differences
#> # A tibble: 1 x 3
#>   Measure lb_stat lb_pvalue
#>   <fct>     <dbl>     <dbl>
#> 1 CO2       7099.         0
#> Ljung-Box test on (difference(count, 12))
#> # A tibble: 1 x 3
#>   Measure lb_stat lb_pvalue
#>   <fct>     <dbl>     <dbl>
#> 1 CO2       3081.         0
#> Ljung-Box test on (difference(count, 12) + difference())
#> # A tibble: 1 x 3
#>   Measure lb_stat lb_pvalue
#>   <fct>     <dbl>     <dbl>
#> 1 CO2        82.7  1.48e-13
```

1.1.2 Unitroot KPSS Test - fix number of seasonal differences/differences required

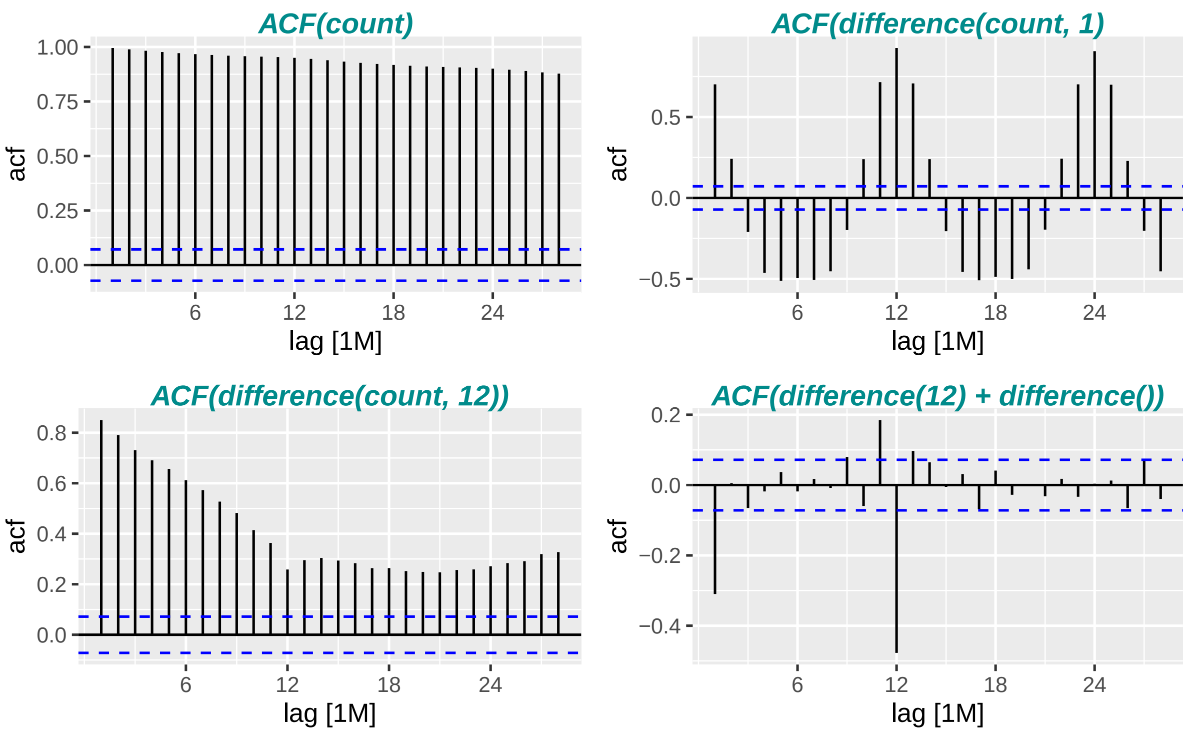
kpss test of stationary
Null Hypothesis of stationary in a given time series
=> H_0 to be rejected for $p < \alpha = 0.05$

unitroot_nsdiffs/ndiff provides minimum number of seasonal differences/differences required for a stationary series. First fix required seasonal differences and then apply ndiffs to the seasonally differenced data.

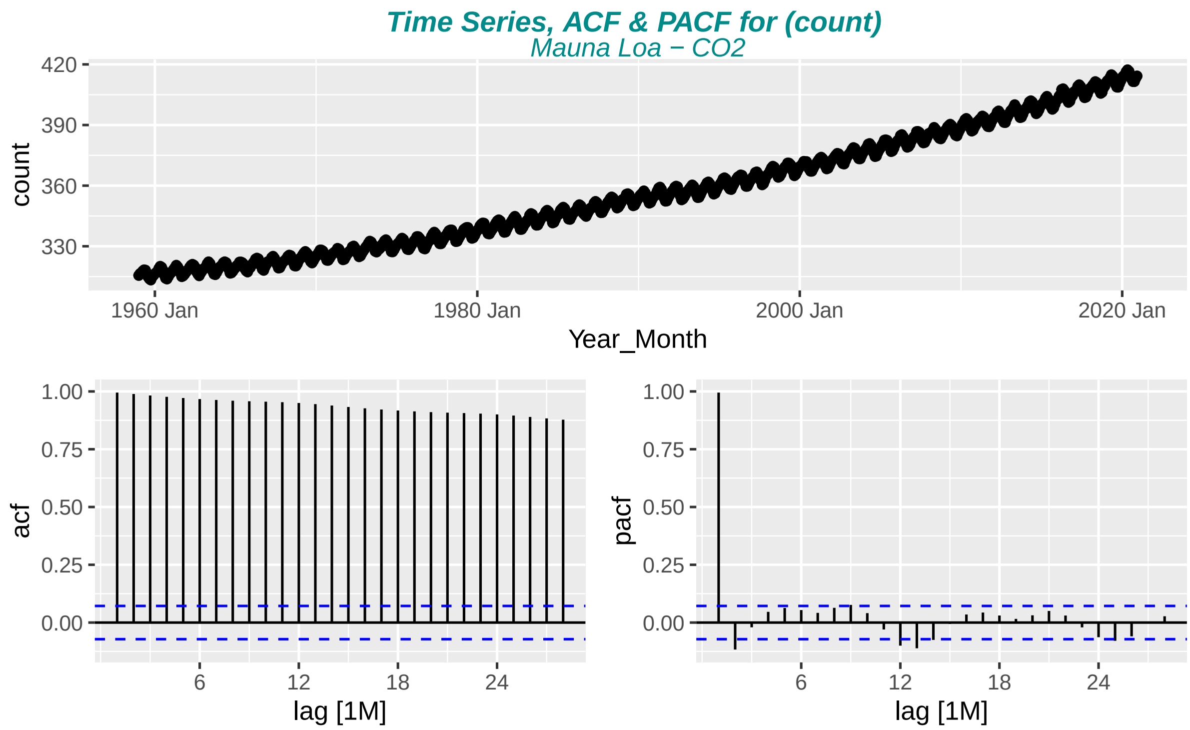
- returns 1 => for stationarity one seasonal difference rsp. difference is required

```
#> ndiffs gives the number of differences required rsp.
#>  nsdiffs gives the number of seasonal differences required to make
#>  a series stationary (test is based on the KPSS test
#> kpss test, nsdiffs & ndiffs on (count), w/o differences
#> # A tibble: 1 x 5
#>   Measure kpss_stat kpss_pvalue nsdiffs ndiffs
#>   <fct>       <dbl>       <dbl>   <int>  <int>
#> 1 CO2          10.6        0.01       1      1
#> kpss test, nsdiffs & ndiffs on  (difference(count, 12)
#> # A tibble: 1 x 5
#>   Measure kpss_stat kpss_pvalue nsdiffs ndiffs
#>   <fct>       <dbl>       <dbl>   <int>  <int>
#> 1 CO2          5.29        0.01       0      1
#> kpss test, nsdiffs & ndiffs on (difference(count, 12) %>% difference(1))
#> # A tibble: 1 x 5
#>   Measure kpss_stat kpss_pvalue nsdiffs ndiffs
#>   <fct>       <dbl>       <dbl>   <int>  <int>
#> 1 CO2       0.00690         0.1       0      0
```

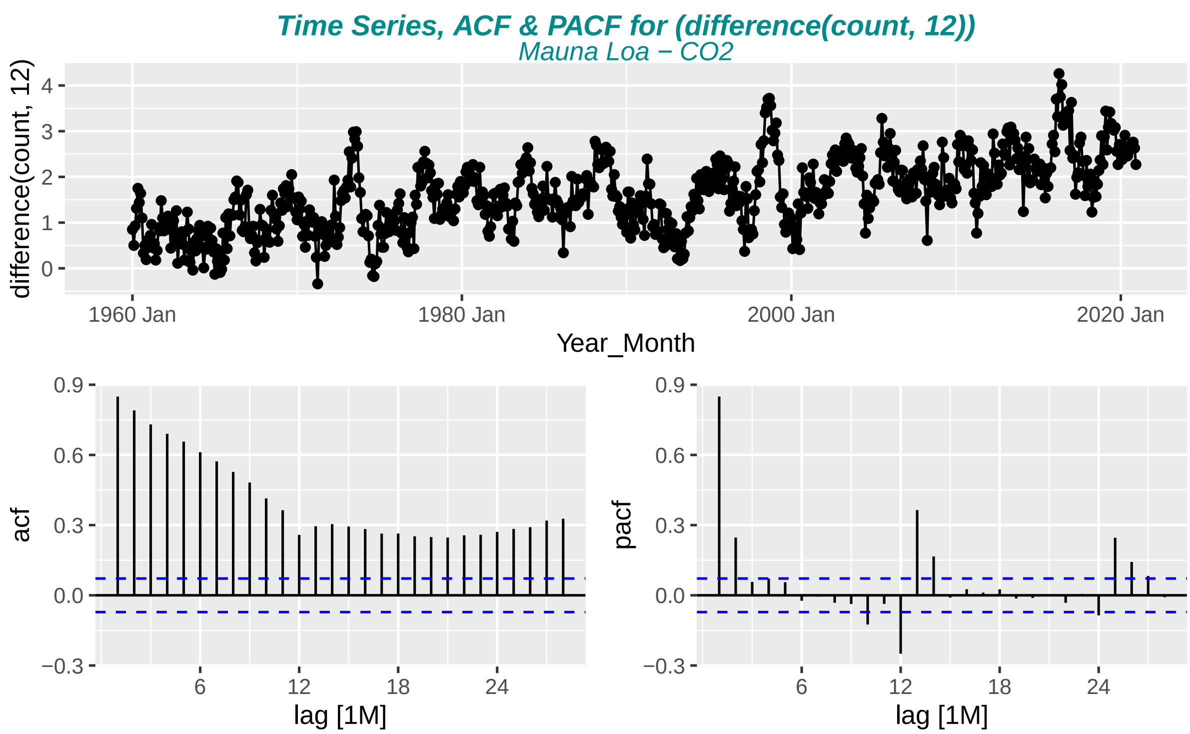
1.1.3 ACF Plots of Differences



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

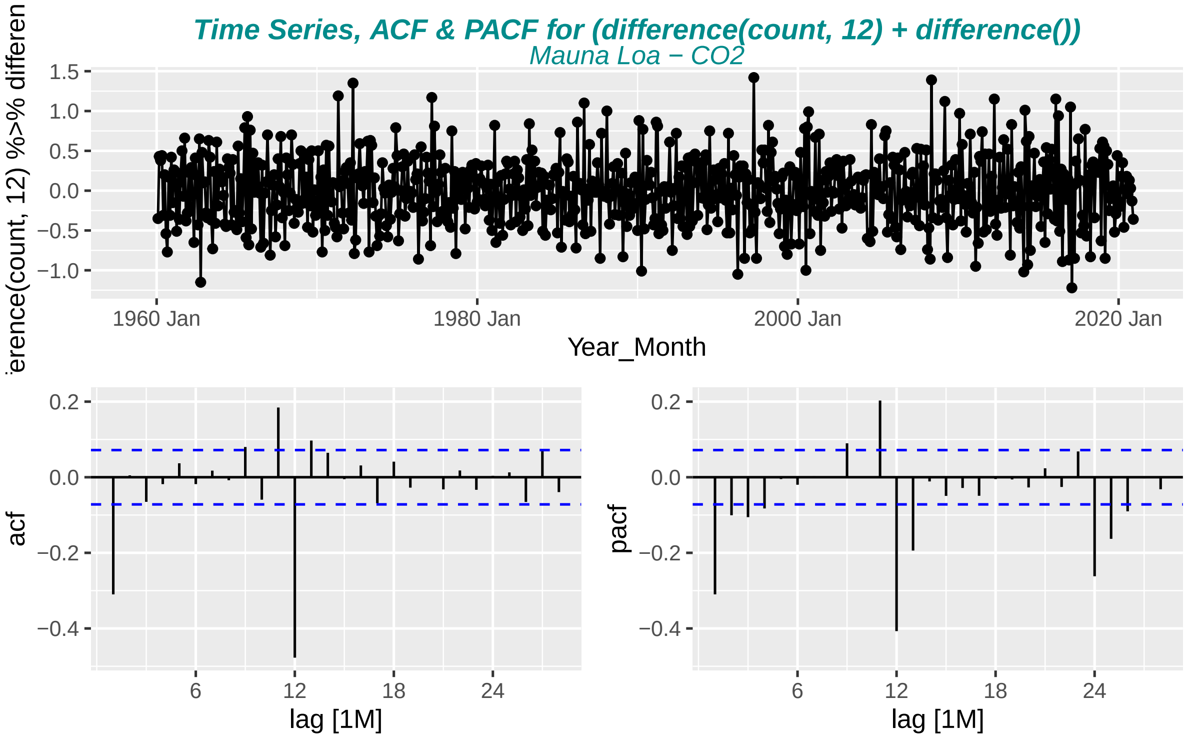



```
#> # A tibble: 1 x 2
#>     Sum  Mean
#>   <dbl> <dbl>
#> 1 1179.  1.61
```




```
#> # A tibble: 1 x 2
#>     Sum  Mean
```

```
#>    <dbl> <dbl>
#> 1 1179.   1.61
```




```
#> # A tibble: 1 x 2
#>     Sum    Mean
#>   <dbl>   <dbl>
#> 1  1.42 0.00194
```

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 togehter). 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 togehter). Best Forecast accuracy is with ETS(A,A,A), ETS(A,Ad,A), ETS(A,N,A),

Trend term "N" for Basel Temperature/Precipitation correspondends to a "pure" exponential smooothing 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 intervall 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" exponentiell smoothing without trend (see above).

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

Seasonal term: either none ("N"), additive ("A") or multiplicative ("M").

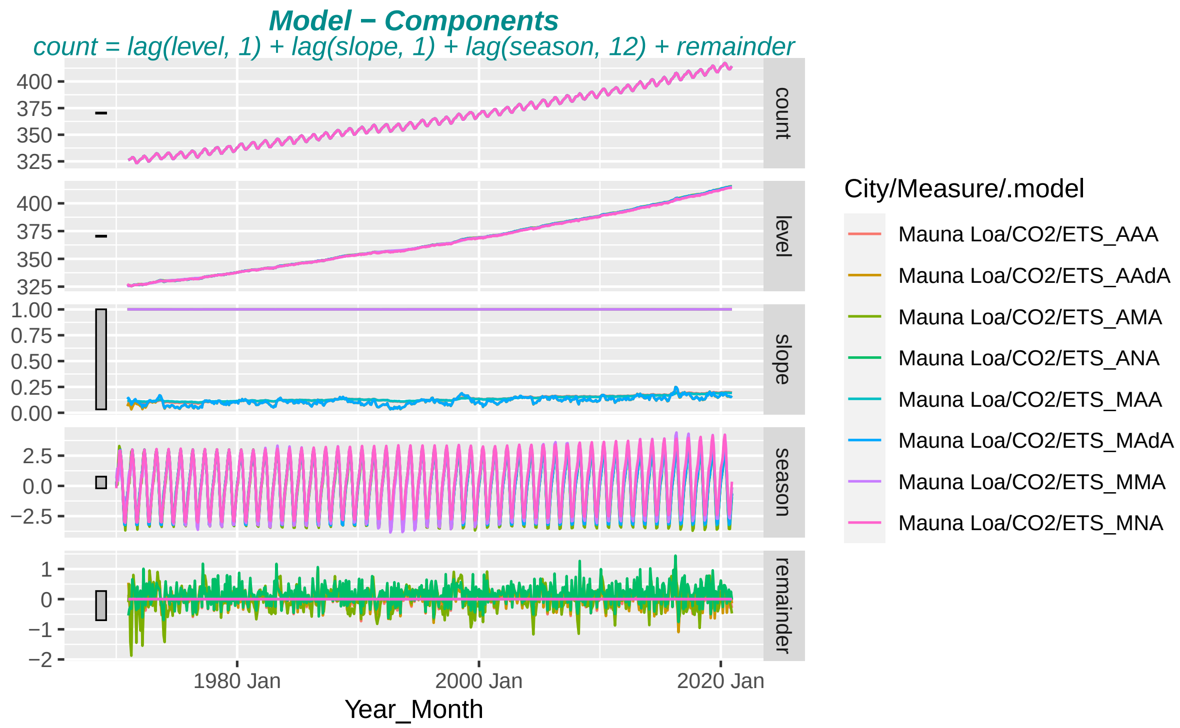
For CO2 and Temperature Data we have a clear seasonal pattern and seasonal term adds always a (more or less) fix amount on level and trend component. Therefore "A" additve term is chosen. For Precipitation the seasonal patttern is only slight. Indead, a multiplicative seasonal term results in "exploding" forecasts.

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

2.1 ETS Models and their componentes

```
#> [1] "model(ETS(count)) => provides best automatically chosen model"
#> # A tibble: 1 x 11
#>   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~ CO2     ETS(co~  7.54e-7  -1220. 2474. 2475. 2549. 0.0981 0.140 6.70e-4
#> Series: count
#> Model: ETS(M,A,A)
#>   Smoothing parameters:
#>     alpha = 0.5957971
#>     beta  = 0.006191047
#>     gamma = 0.08087914
#>
#>   Initial states:
#>         l         b         s1        s2        s3        s4      s5        s6
#>  325.8454 0.1146508 -0.8576564 -1.942806 -3.101869 -3.106011 -1.3612 0.6760767
#>        s7       s8       s9      s10       s11        s12
#>  2.143165 2.902583 2.441888 1.452089 0.7266103 0.02713021
#>
#>   sigma^2:  0
#>
#>      AIC     AICc      BIC
#> 2474.311 2475.362 2549.058
#> Model Selection by Information Criterion - lowest AIC, AICc, BIC
```

```
#> # A tibble: 8 x 11
#>   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 ~ CO2     ETS_M~  7.54e-7  -1220. 2474. 2475. 2549. 0.0981 0.140 6.70e-4
#> 2 Mauna ~ CO2     ETS_A~  1.01e-1  -1224. 2481. 2482. 2556. 0.0984 0.139 2.45e-1
#> 3 Mauna ~ CO2     ETS_M~  8.09e-7  -1241. 2518. 2519. 2597. 0.105  0.164 6.96e-4
#> 4 Mauna ~ CO2     ETS_A~  1.08e-1  -1244. 2524. 2525. 2603. 0.105  0.165 2.54e-1
#> 5 Mauna ~ CO2     ETS_A~  1.39e-1  -1319. 2669. 2670. 2735. 0.136  0.270 2.92e-1
#> 6 Mauna ~ CO2     ETS_M~  1.07e-6  -1326. 2683. 2684. 2749. 0.141  0.281 8.15e-4
#> 7 Mauna ~ CO2     ETS_A~  1.57e-1  -1356. 2746. 2747. 2820. 0.153  0.193 2.97e-1
#> 8 Mauna ~ CO2     ETS_M~  1.80e-6  -1481. 2995. 2996. 3070. 0.224  0.251 1.01e-3
```



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

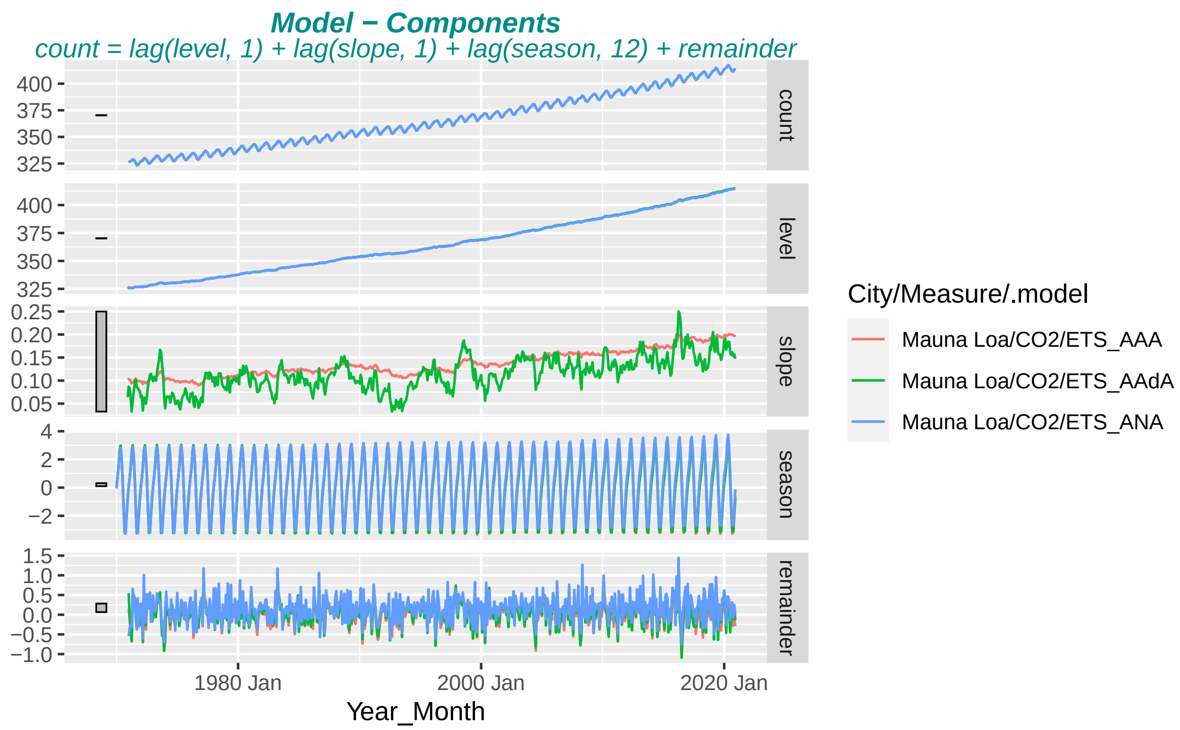
- MAE (Mean absolute error, will lead to forecasts of the median) and
- RMSE (Root mean squared error, lead to forecasts of the mean)

```
#> # A tibble: 8 x 11
#>   City  Measure .model .type        ME  RMSE   MAE      MPE   MAPE  MASE     ACF1
#>   <chr> <fct>   <chr>  <chr>     <dbl> <dbl> <dbl>    <dbl>  <dbl> <dbl>    <dbl>
#> 1 Maun~ CO2     ETS_M~ Trai~  0.0204   0.313 0.245  0.00523 0.0670 0.136   0.0668
#> 2 Maun~ CO2     ETS_A~ Trai~  0.0201   0.314 0.245  0.00527 0.0672 0.137   0.0692
#> 3 Maun~ CO2     ETS_M~ Trai~  0.0546   0.324 0.254  0.0147  0.0696 0.142   0.0274
#> 4 Maun~ CO2     ETS_A~ Trai~  0.0544   0.325 0.254  0.0147  0.0696 0.142   0.00158
#> 5 Maun~ CO2     ETS_A~ Trai~  0.158    0.368 0.292  0.0427  0.0799 0.163  -0.192
#> 6 Maun~ CO2     ETS_M~ Trai~  0.164    0.375 0.298  0.0444  0.0815 0.166  -0.137
#> 7 Maun~ CO2     ETS_A~ Trai~ -0.00653  0.391 0.297 -0.00229 0.0820 0.165   0.416
#> 8 Maun~ CO2     ETS_M~ Trai~  0.0362   0.473 0.365  0.00915 0.101  0.203   0.580
```

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

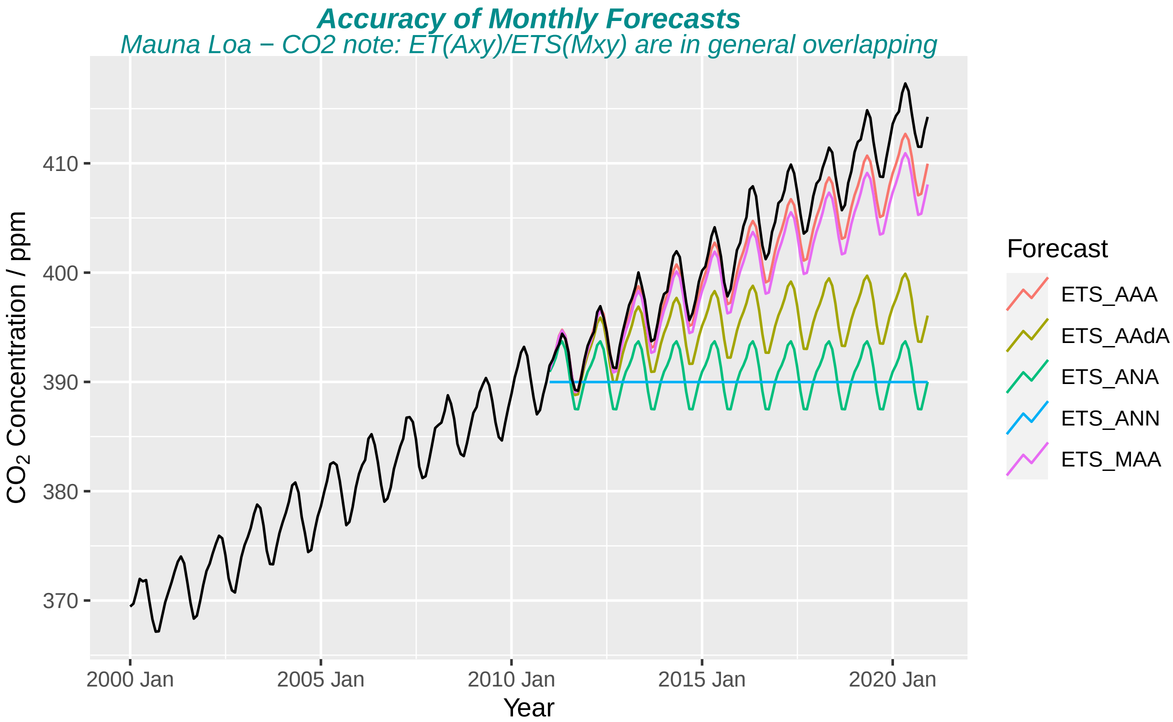
```
#> Null Hypothesis of independence/white noise for residuals - for p < 0.05: reject H_0
#> # A tibble: 8 x 5
#>   City      Measure .model   lb_stat lb_pvalue
#>   <chr>     <fct>   <chr>      <dbl>     <dbl>
#> 1 Mauna Loa CO2     ETS_AAA     41.3 0.0814
#> 2 Mauna Loa CO2     ETS_MAA     46.0 0.0311
#> 3 Mauna Loa CO2     ETS_MNA     59.7 0.00101
#> 4 Mauna Loa CO2     ETS_AAdA    64.3 0.000273
#> 5 Mauna Loa CO2     ETS_MAdA    64.9 0.000224
#> 6 Mauna Loa CO2     ETS_ANA     68.7 0.0000721
#> 7 Mauna Loa CO2     ETS_AMA    253.  0
#> 8 Mauna Loa CO2     ETS_MMA    692.  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

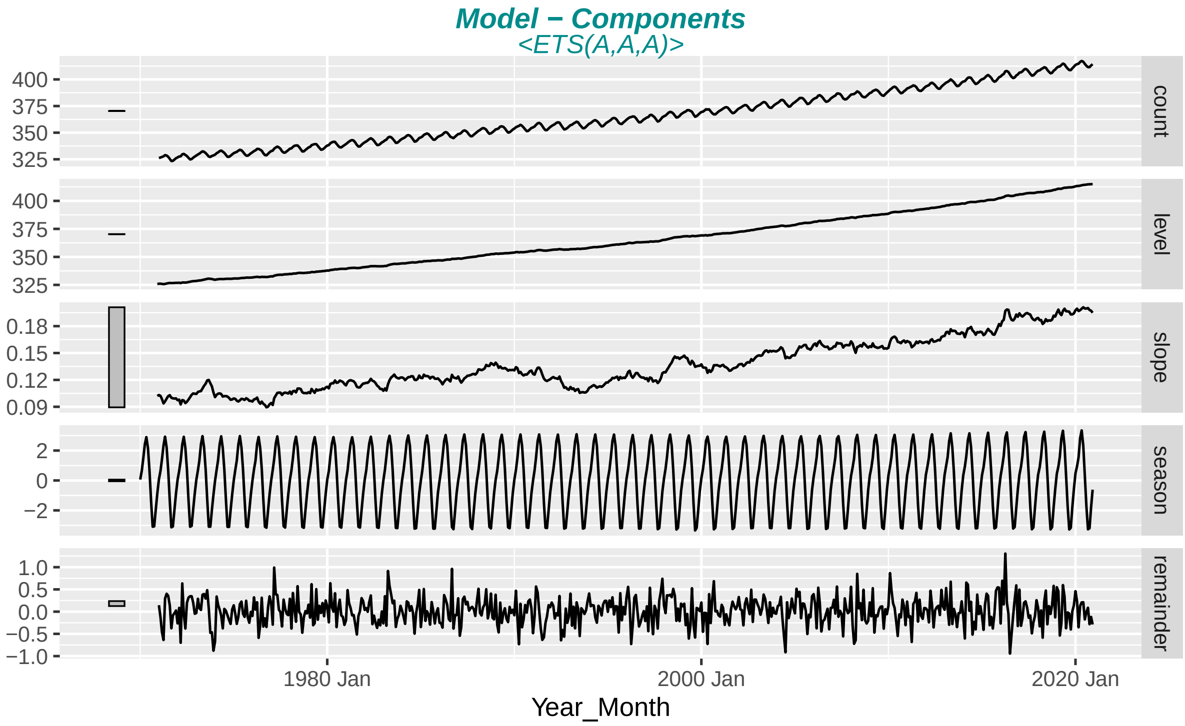
```
#> # A tibble: 5 x 11
#>   .model   City      Measure .type    ME  RMSE   MAE   MPE  MAPE  MASE  ACF1
#>   <chr>    <chr>     <fct>   <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
#> 1 ETS_AAA  Mauna Loa CO2     Test   1.88  2.41  1.93 0.460 0.473  1.18 0.953
#> 2 ETS_MAA  Mauna Loa CO2     Test   2.82  3.48  2.84 0.690 0.698  1.74 0.964
#> 3 ETS_AAdA Mauna Loa CO2     Test   7.75  9.60  7.77 1.90  1.90   4.74 0.976
#> 4 ETS_ANA  Mauna Loa CO2     Test  12.1  14.1  12.1  2.97  2.97   7.38 0.976
#> 5 ETS_ANN  Mauna Loa CO2     Test  12.8  14.8  12.9  3.15  3.16   7.86 0.964
```

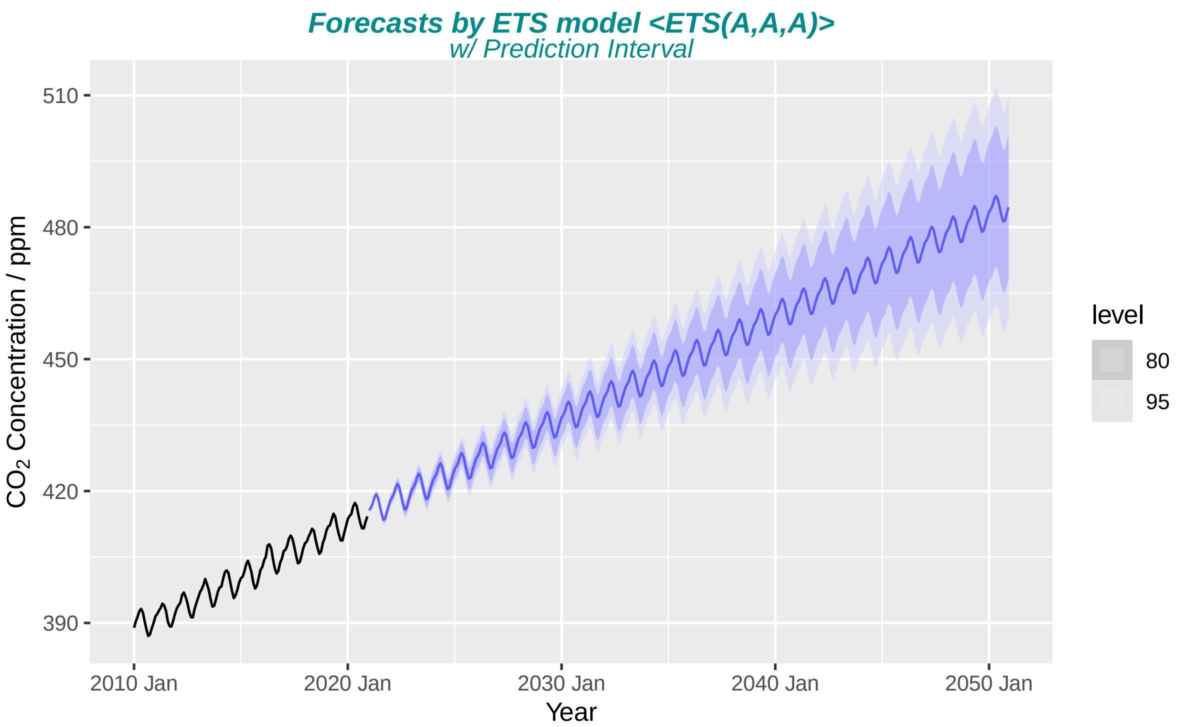


2.2 Forecasting with selected ETS model <ETS(A,A,A)>

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 = 0.5800712
#>     beta  = 0.00770667
#>     gamma = 0.1044484
#>
#>   Initial states:
#>         l          b        s1        s2        s3        s4        s5        s6
#>   325.8617 0.1020527 -0.860074 -1.930963 -3.068509 -3.095545 -1.358178 0.6678765
#>        s7       s8       s9      s10     s11        s12
#>   2.164707 2.898347 2.416044 1.44868 0.65598 0.06163514
#>
#>   sigma^2:  0.1011
#>
#>      AIC     AICc      BIC
#> 2481.057 2482.108 2555.805
```

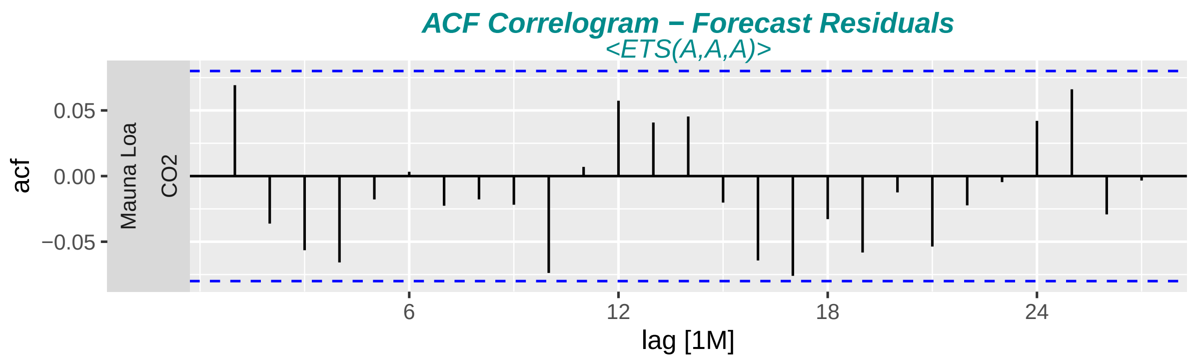


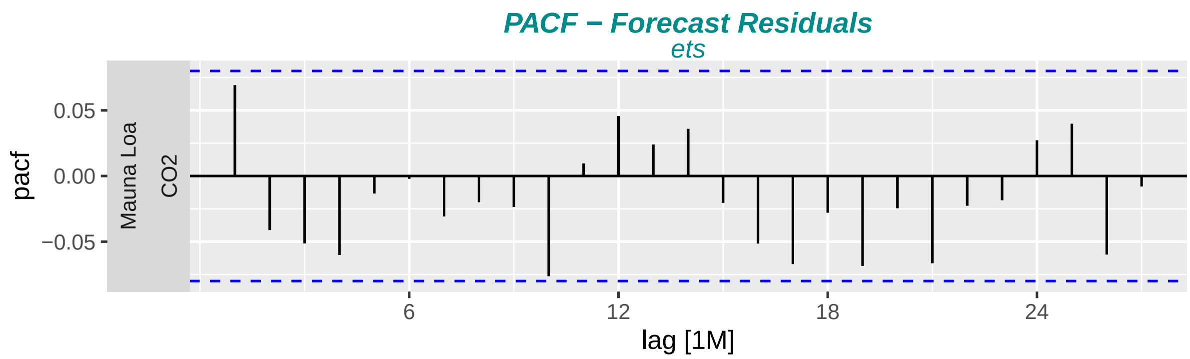


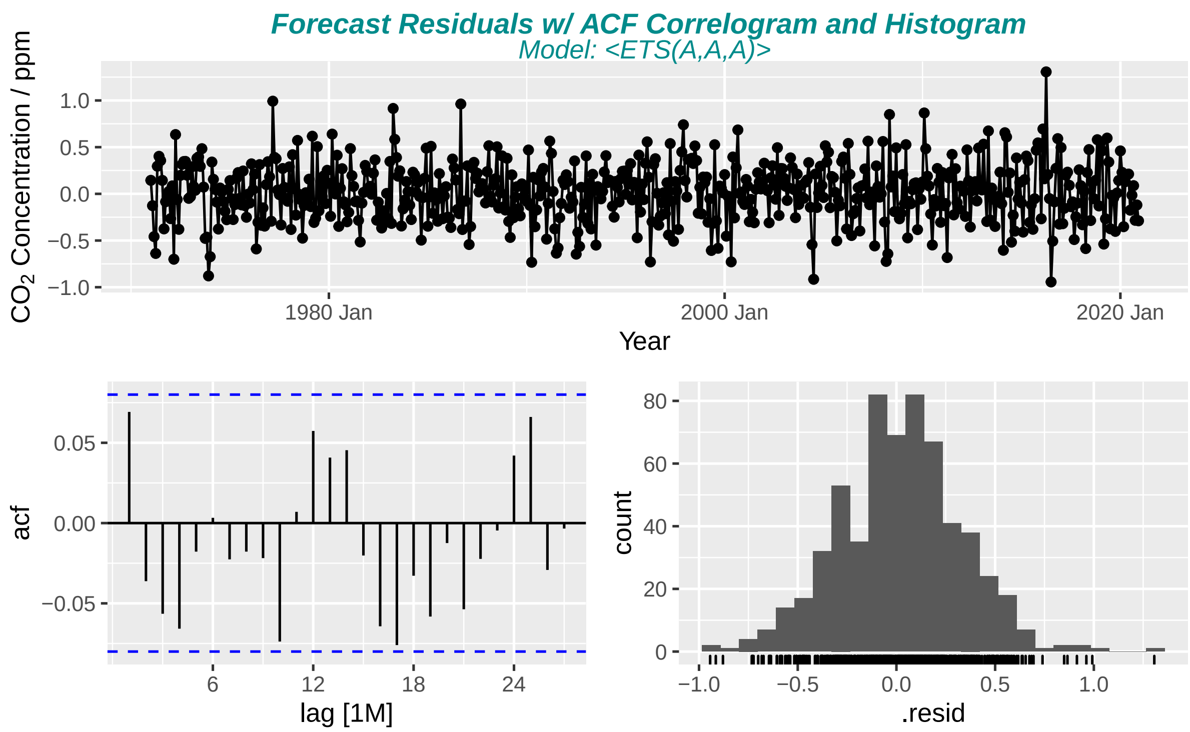
2.2.2 Residual Stationarity

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

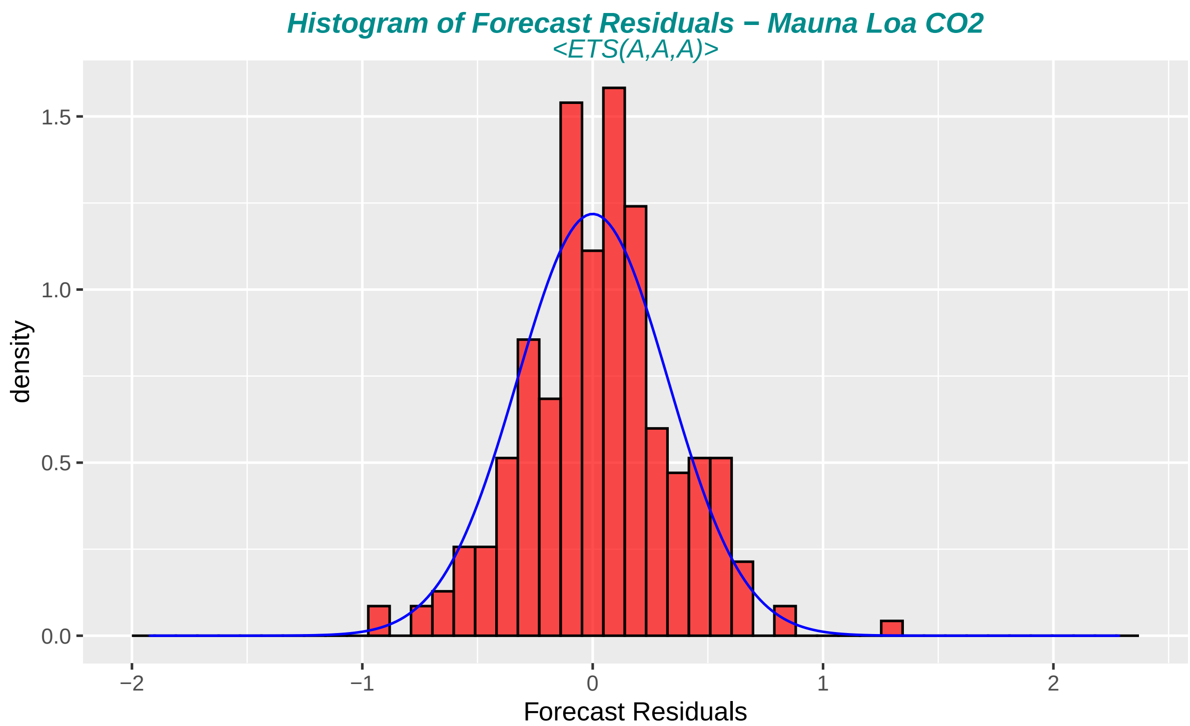






2.2.3 Histogram of forecast residuals with overlaid normal curve

```
#> 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       34.9     0.246
```



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

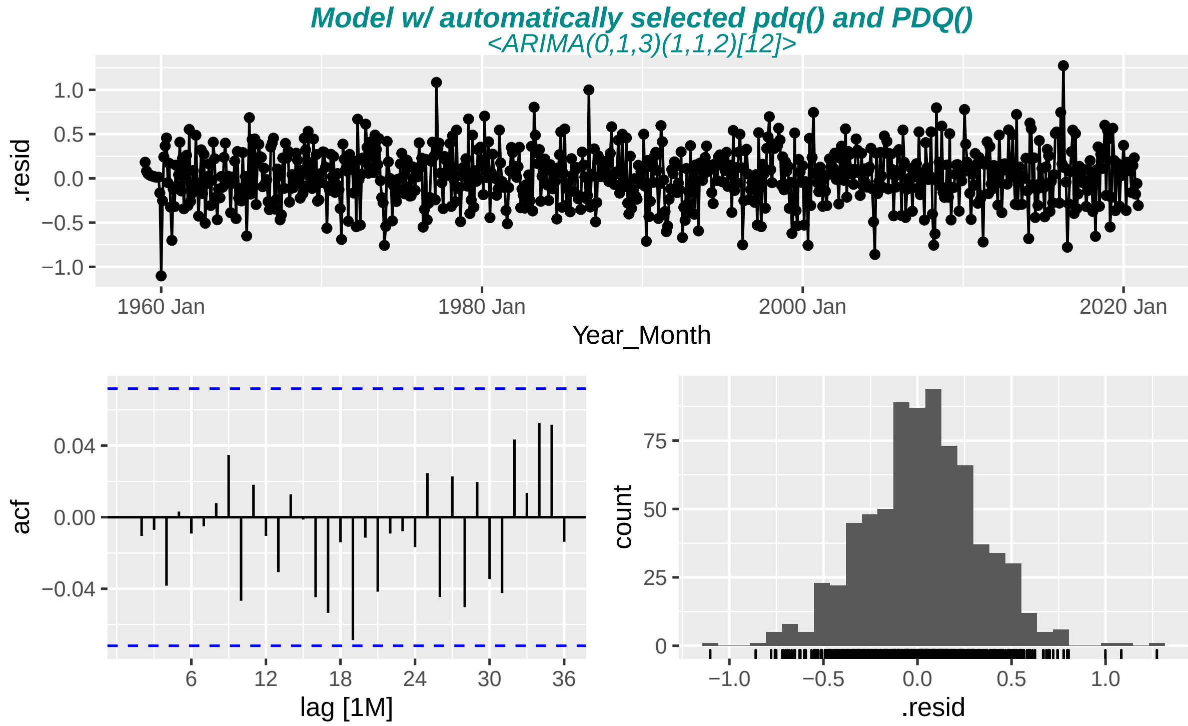
Non-seasonal ARIMA models are generally denoted ARIMA(p,d,q) where parameters p, d, and q are non-negative integers, * p is the order (number of time lags) of the autoregressive model * d is the degree of differencing (number of times the data have had past values subtracted) * q is the order of the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

```
#> # A tibble: 1 x 10
#>   City       Measure .model sigma2 log_lik   AIC  AICc   BIC ar_roots   ma_roots
#>   <chr>      <fct>   <chr>   <dbl>   <dbl> <dbl> <dbl> <dbl> <list>     <list>
#> 1 Mauna Loa CO2     arima  0.0958   -178.  371.  371.  403. <cpl [12]> <cpl [27~
#> Series: count
#> Model: ARIMA(0,1,3)(1,1,2)[12]
```

```
#>
#> Coefficients:
#>           ma1      ma2      ma3     sar1     sma1     sma2
#>       -0.3580  -0.0318  -0.0579  -0.3758  -0.4758  -0.3377
#> s.e.   0.0375   0.0403   0.0371   1.4777   1.4780   1.2764
#>
#> sigma^2 estimated as 0.09581:  log likelihood=-178.41
#> AIC=370.83   AICc=370.98   BIC=402.99
```




```
#> Model Selection by Information Criterion - lowest AIC, AICc, BIC
#>  choose p, q parameter accordingly - but only for same d, D values
#> # A tibble: 12 x 10
#>    City    Measure .model    sigma2 log_lik   AIC  AICc   BIC ar_roots   ma_roots
#>    <chr>   <fct>   <chr>      <dbl>   <dbl> <dbl> <dbl> <dbl> <list>     <list>
#>  1 Mauna ~ CO2     ARIMA_11~  0.100   -157.  324.  324.  346. <cpl [1]> <cpl [2~
#>  2 Mauna ~ CO2     ARIMA_01~  0.100   -157.  325.  325.  346. <cpl [0]> <cpl [2~
#>  3 Mauna ~ CO2     ARIMA_21~  0.100   -158.  325.  325.  347. <cpl [2]> <cpl [1~
#>  4 Mauna ~ CO2     ARIMA_11~  0.101   -158.  327.  327.  353. <cpl [13~ <cpl [2~
#>  5 Mauna ~ CO2     ARIMA_10~  0.127   -223.  455.  455.  477. <cpl [14~ <cpl [0~
#>  6 Mauna ~ CO2     ARIMA_20~  0.127   -223.  455.  455.  477. <cpl [14~ <cpl [0~
#>  7 Mauna ~ CO2     ARIMA_21~  0.128   -225.  458.  458.  476. <cpl [14~ <cpl [0~
#>  8 Mauna ~ CO2     ARIMA_10~  0.130   -230.  471.  471.  493. <cpl [25~ <cpl [0~
#>  9 Mauna ~ CO2     ARIMA_01~  0.143   -259.  522.  522.  531. <cpl [12~ <cpl [0~
#> 10 Mauna ~ CO2     ARIMA_11~  0.167   -303.  612.  612.  625. <cpl [1]> <cpl [1~
#> 11 Mauna ~ CO2     ARIMA_01~  0.167   -303.  612.  612.  625. <cpl [0]> <cpl [2~
#> 12 Mauna ~ CO2     ARIMA_11~  0.171   -310.  623.  623.  632. <cpl [1]> <cpl [0~
```

Good models are obtained by minimising the AIC, AICc or BIC (see glance(fit_arima) output). The preference is to use the AICc to selec 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)

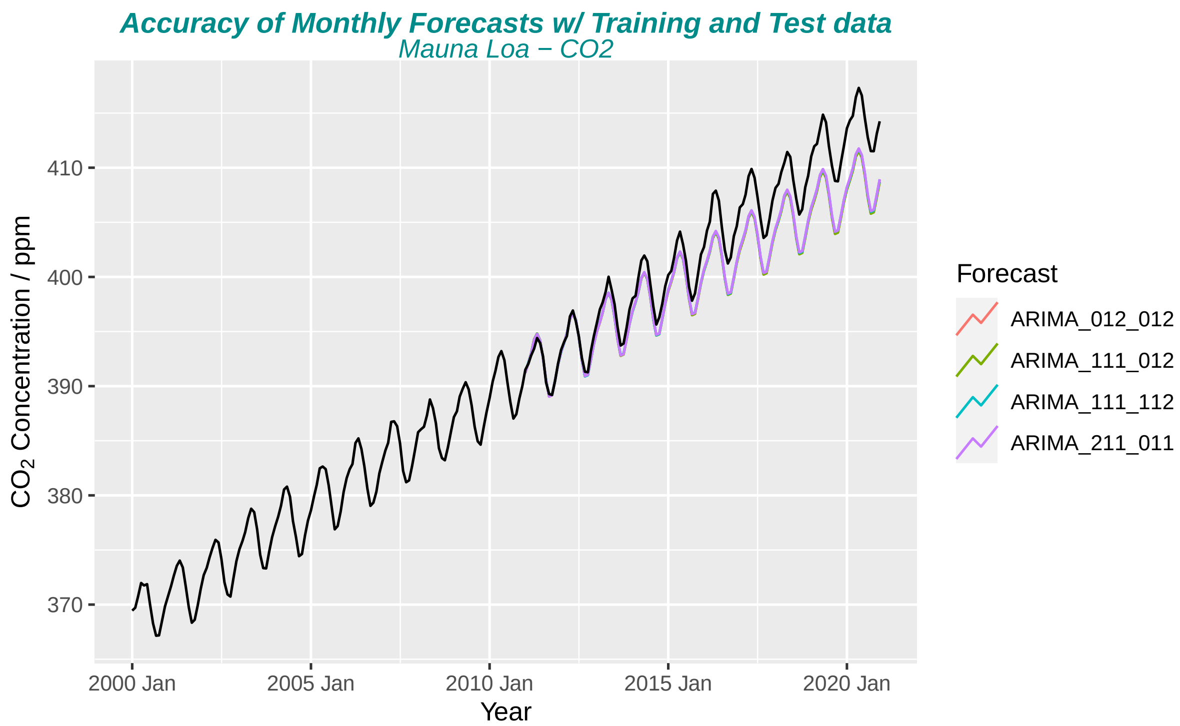
```
#> # A tibble: 14 x 11
#>    City  Measure .model .type        ME    RMSE     MAE       MPE     MAPE
#>    <chr> <fct>   <chr>  <chr>     <dbl>   <dbl>   <dbl>     <dbl>    <dbl>
#>  1 Maun~ CO2     ARIMA~ Trai~   0.0245    0.312   0.241   6.50e-3   0.0663
#>  2 Maun~ CO2     ARIMA~ Trai~   0.0240    0.312   0.242   6.39e-3   0.0663
#>  3 Maun~ CO2     ARIMA~ Trai~   0.0245    0.312   0.242   6.52e-3   0.0663
#>  4 Maun~ CO2     ARIMA~ Trai~   0.0243    0.312   0.242   6.45e-3   0.0663
#>  5 Maun~ CO2     ARIMA~ Trai~   0.00910   0.351   0.280   1.92e-3   0.0769
#>  6 Maun~ CO2     ARIMA~ Trai~   0.00910   0.351   0.280   1.92e-3   0.0769
#>  7 Maun~ CO2     ARIMA~ Trai~   0.00492   0.353   0.279   1.32e-3   0.0766
#>  8 Maun~ CO2     ARIMA~ Trai~   0.00857   0.356   0.280   1.56e-3   0.0767
#>  9 Maun~ CO2     ARIMA~ Trai~   0.00290   0.374   0.289   7.65e-4   0.0790
#> 10 Maun~ CO2     ARIMA~ Trai~   0.00362   0.404   0.313   9.71e-4   0.0858
#> 11 Maun~ CO2     ARIMA~ Trai~   0.00359   0.404   0.313   9.63e-4   0.0859
#> 12 Maun~ CO2     ARIMA~ Trai~   0.00245   0.408   0.320   6.50e-4   0.0877
#> 13 Maun~ CO2     ARIMA~ Trai~ NaN       NaN     NaN     NaN       NaN
#> 14 Maun~ CO2     ARIMA~ Trai~ NaN       NaN     NaN     NaN       NaN
#> # ... with 2 more variables: MASE <dbl>, ACF1 <dbl>
```

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: 14 x 5
#>    City      Measure .model        lb_stat lb_pvalue
#>    <chr>     <fct>   <chr>           <dbl>     <dbl>
#>  1 Mauna Loa CO2     ARIMA_111_012    25.5  7.02e- 1
#>  2 Mauna Loa CO2     ARIMA_012_012    25.7  6.89e- 1
#>  3 Mauna Loa CO2     ARIMA_211_011    26.1  6.69e- 1
#>  4 Mauna Loa CO2     ARIMA_111_112    26.5  6.48e- 1
#>  5 Mauna Loa CO2     ARIMA_100_110    71.2  3.35e- 5
#>  6 Mauna Loa CO2     ARIMA_200_110    71.2  3.35e- 5
#>  7 Mauna Loa CO2     ARIMA_210_110    82.4  8.73e- 7
#>  8 Mauna Loa CO2     ARIMA_100_210   113.   1.34e-11
#>  9 Mauna Loa CO2     ARIMA_010_110   152.   0.
#> 10 Mauna Loa CO2     ARIMA_012_010   164.   0.
#> 11 Mauna Loa CO2     ARIMA_110_010   180.   0.
#> 12 Mauna Loa CO2     ARIMA_111_010   164.   0.
#> 13 Mauna Loa CO2     ARIMA_002_200    NA   NA
#> 14 Mauna Loa CO2     ARIMA_301_200    NA   NA
```

3.1.3 Forecast Accuracy with Training/Test Data

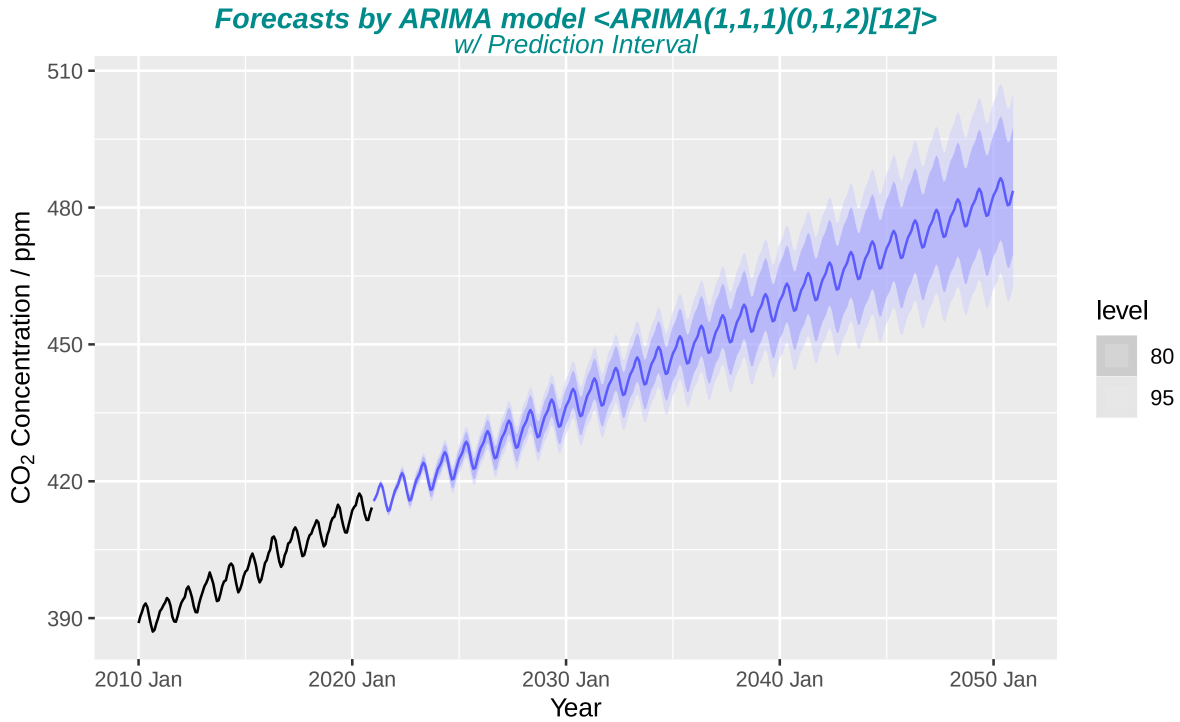
```
#> # A tibble: 4 x 11
#>   .model        City     Measure .type    ME  RMSE   MAE   MPE  MAPE  MASE  ACF1
#>   <chr>         <chr>    <fct>   <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
#> 1 ARIMA_211_011 Mauna L~ CO2     Test   2.39  2.99  2.43 0.587 0.596  1.49 0.961
#> 2 ARIMA_111_112 Mauna L~ CO2     Test   2.41  3.02  2.45 0.591 0.601  1.50 0.962
#> 3 ARIMA_012_012 Mauna L~ CO2     Test   2.49  3.11  2.53 0.610 0.621  1.55 0.963
#> 4 ARIMA_111_012 Mauna L~ CO2     Test   2.50  3.13  2.54 0.613 0.624  1.55 0.963
```



3.2 CO2 - Forecasting with selected ARIMA model <ARIMA(1,1,1)(0,1,2)[12]>

3.2.1 Forecast Plot of selected ARIMA model

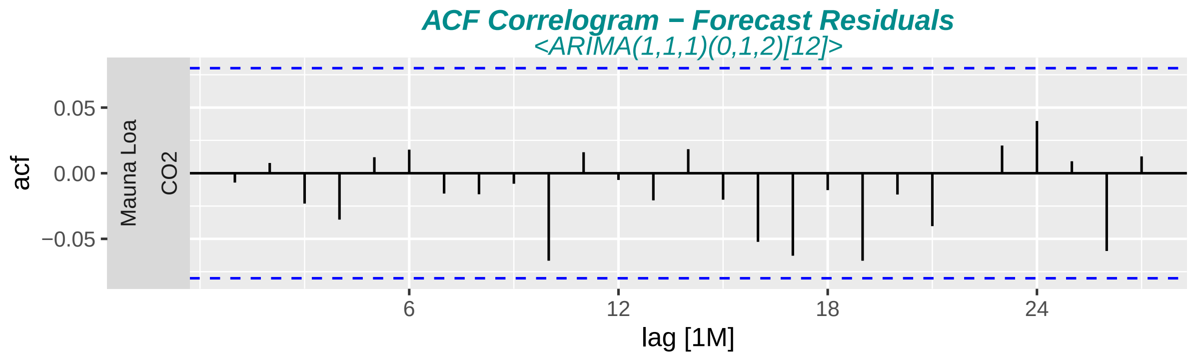
```
#> Provide model coefficients by report(fit_model)
#> Series: count
#> Model: ARIMA(1,1,1)(0,1,2)[12]
#>
#> Coefficients:
#>          ar1      ma1     sma1     sma2
#>       0.1743  -0.5198  -0.8330  -0.0393
#> s.e.  0.1108   0.0966   0.0409   0.0396
#>
#> sigma^2 estimated as 0.1002:  log likelihood=-157.14
#> AIC=324.28   AICc=324.39   BIC=346.16
```

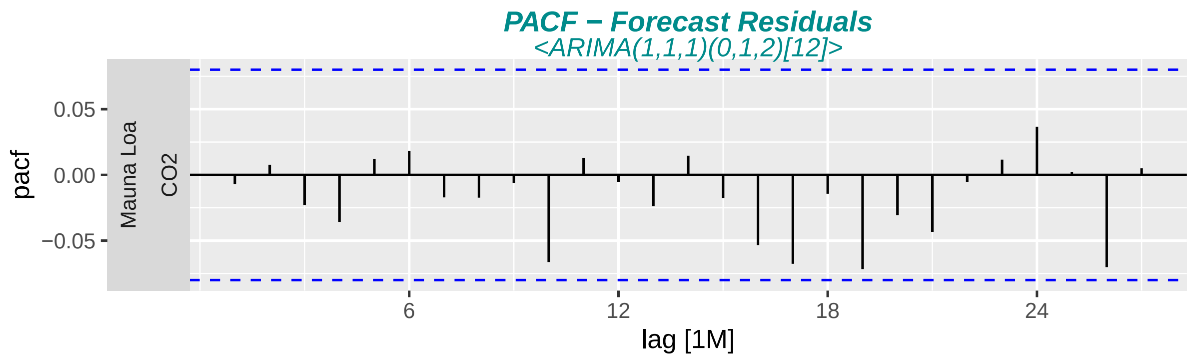


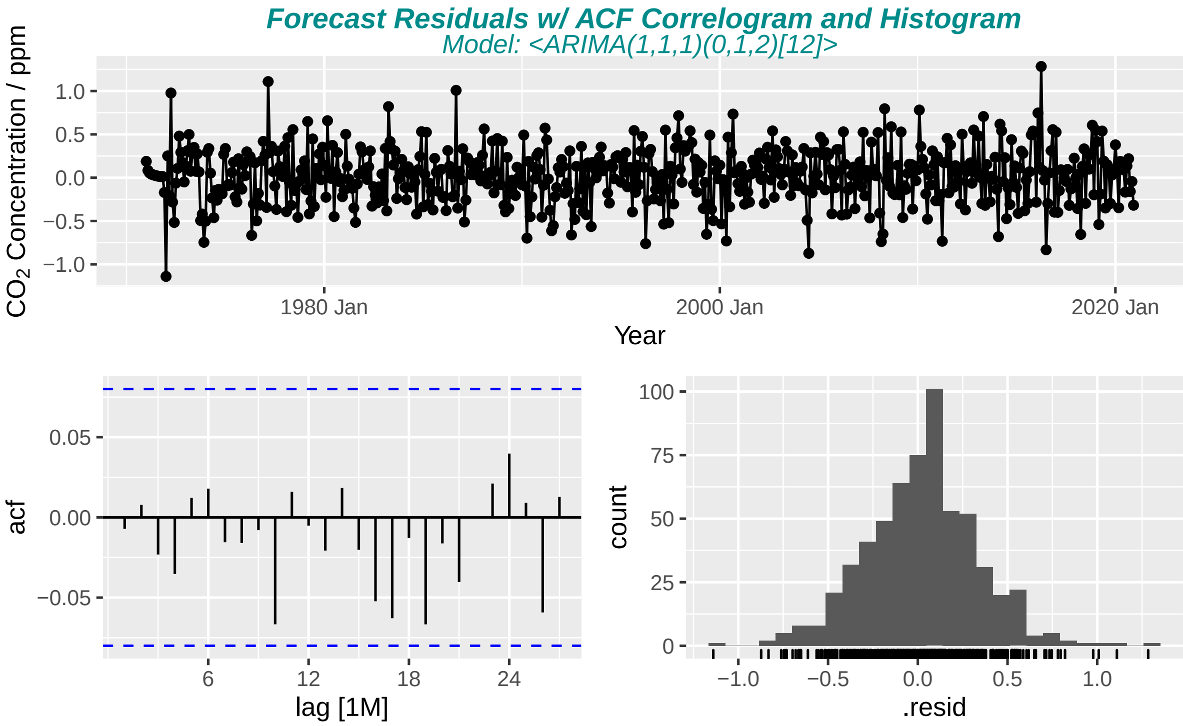
3.2.2 Residual Stationarity

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

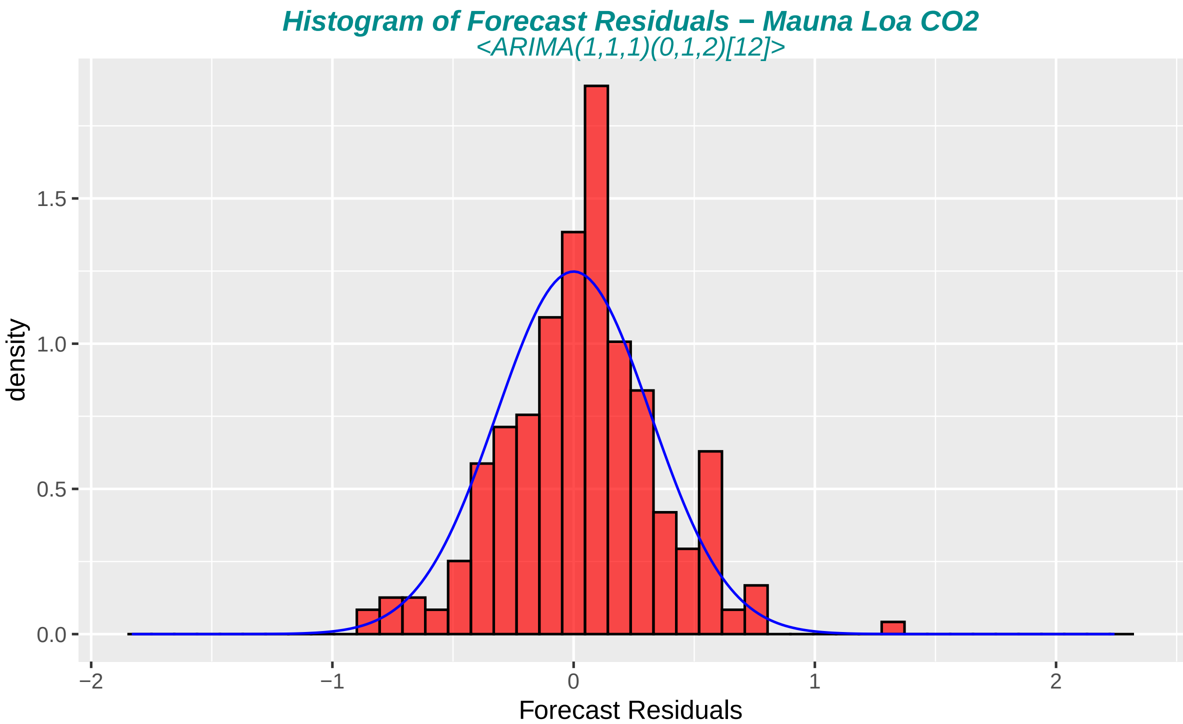






3.2.3 Histogram of forecast residuals with overlaid normal curve

```
#> 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     22.0     0.855
```



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 rsp. ARIMA model the RMSE / MAE values. Lower values indicate a more accurate model based on the test set RMSE, . . . , MASE.

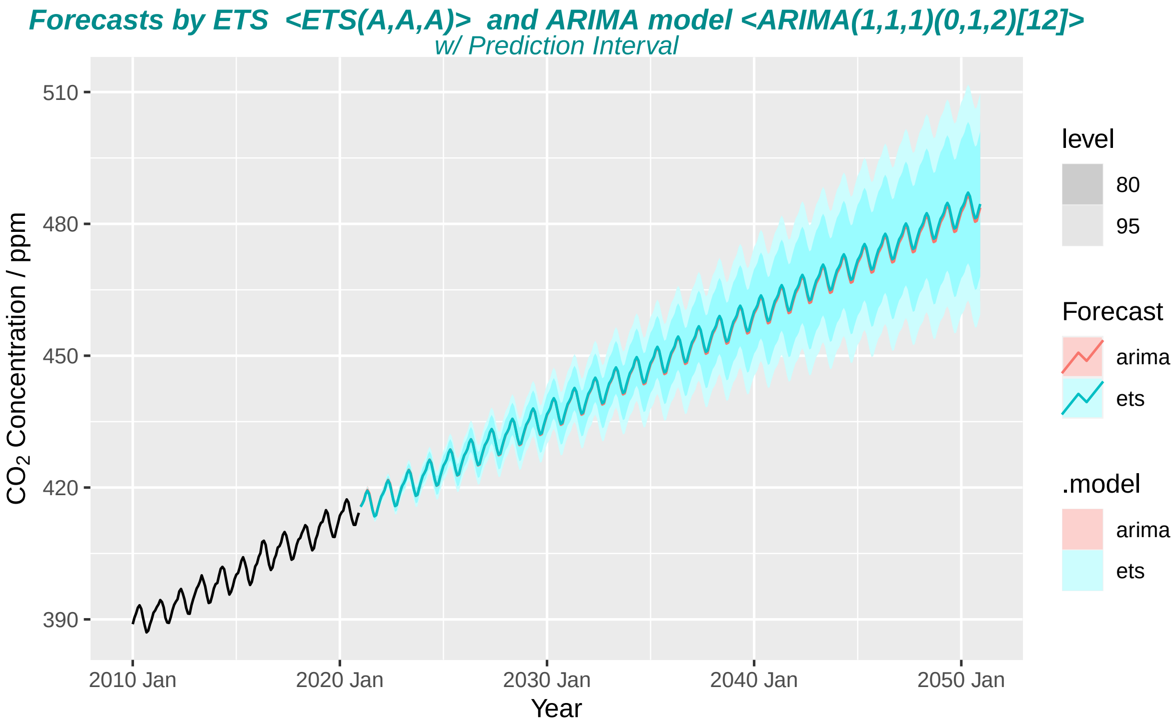
- Residual Accuracy with one-step-ahead fitted residuals
- Forecast Accuracy with Training/Test Data

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

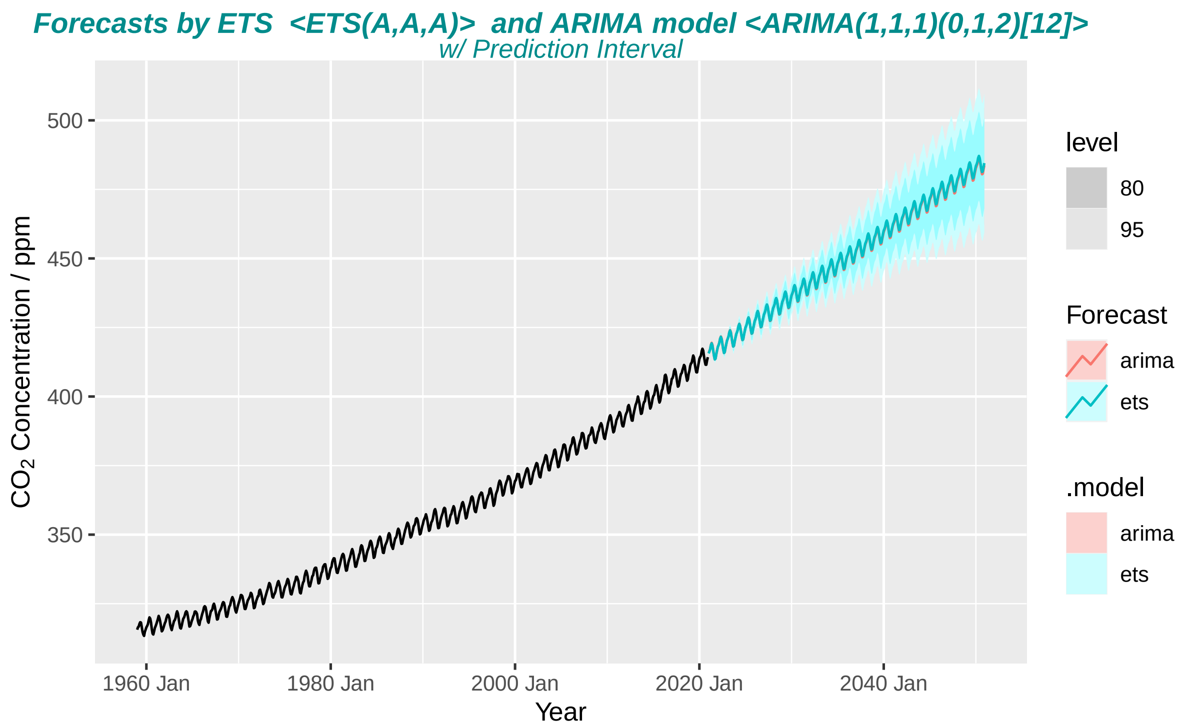
```
#> # A tibble: 4 x 11
#>   City    Measure .model    .type     ME  RMSE   MAE     MPE   MAPE  MASE     ACF1
#>   <chr>   <fct>   <chr>     <chr>  <dbl> <dbl> <dbl>   <dbl>  <dbl> <dbl>    <dbl>
#> 1 Mauna~ CO2     ets       Trai~ 0.0201 0.314 0.245 0.00527 0.0672 0.137   0.0692
#> 2 Mauna~ CO2     arima     Trai~ 0.0245 0.312 0.241 0.00650 0.0663 0.134 -0.00712
#> 3 Mauna~ CO2     ETS_AAA   Test  1.88   2.41  1.93  0.460   0.473  1.18    0.953
#> 4 Mauna~ CO2     ARIMA_1~ Test  2.50   3.13  2.54  0.613   0.624  1.55    0.963
```

4.0.2 Forecast Plot of selected ETS and ARIMA model




```
#> # A tsibble: 6 x 8 [1M]
#> # Key:       City, Measure, .model [2]
#> # Groups:    City, Measure, .model [2]
```

```
#>   City  Measure .model Year_Month        count .mean                        `80%`
#>   <chr> <fct>   <chr>       <mth>       <dist> <dbl>                        <hilo>
#> 1 Maun~ CO2     arima    2021 Jan  N(416, 0.1)  416. [415.2749, 416.0862]80
#> 2 Maun~ CO2     arima    2021 Feb N(416, 0.14)  416. [415.9544, 416.9241]80
#> 3 Maun~ CO2     arima    2021 Mrz N(417, 0.18)  417. [416.6710, 417.7540]80
#> 4 Maun~ CO2     ets      2021 Jan  N(416, 0.1)  416. [415.2278, 416.0429]80
#> 5 Maun~ CO2     ets      2021 Feb N(416, 0.14)  416. [415.8290, 416.7744]80
#> 6 Maun~ CO2     ets      2021 Mrz N(417, 0.17)  417. [416.5535, 417.6162]80
#> # ... with 1 more variable: `95%` <hilo>
#> # A tsibble: 6 x 8 [1M]
#> # Key:       City, Measure, .model [2]
#> # Groups:    City, Measure, .model [2]
#>   City  Measure .model Year_Month       count .mean                        `80%`
#>   <chr> <fct>   <chr>       <mth>      <dist> <dbl>                        <hilo>
#> 1 Maun~ CO2     arima    2050 Okt N(481, 116)  481. [466.8886, 494.5019]80
#> 2 Maun~ CO2     arima    2050 Nov N(482, 117)  482. [468.4415, 496.1453]80
#> 3 Maun~ CO2     arima    2050 Dez N(484, 118)  484. [469.7903, 497.5843]80
#> 4 Maun~ CO2     ets      2050 Okt N(482, 163)  482. [465.2054, 497.9013]80
#> 5 Maun~ CO2     ets      2050 Nov N(483, 164)  483. [466.7630, 499.5720]80
#> 6 Maun~ CO2     ets      2050 Dez N(485, 165)  485. [468.0651, 500.9873]80
#> # ... with 1 more variable: `95%` <hilo>
```




```
#> # A tibble: 180 x 5
#> # Groups:   City, Measure, .model, Year [60]
#>    City      Measure .model  Year Year_avg
#>    <chr>     <fct>   <chr>  <dbl>    <dbl>
#>  1 Mauna Loa CO2     arima   2021     416.
#>  2 Mauna Loa CO2     arima   2021     416.
#>  3 Mauna Loa CO2     arima   2021     417.
#>  4 Mauna Loa CO2     arima   2022     418.
#>  5 Mauna Loa CO2     arima   2022     419.
#>  6 Mauna Loa CO2     arima   2022     420.
#>  7 Mauna Loa CO2     arima   2023     420.
```

```
#>  8 Mauna Loa CO2     arima   2023     421.
#>  9 Mauna Loa CO2     arima   2023     422.
#> 10 Mauna Loa CO2     arima   2024     423.
#> # ... with 170 more rows
#> # A tibble: 180 x 5
#> # Groups:   City, Measure, .model, Year [60]
#>    City      Measure .model  Year Year_avg
#>    <chr>     <fct>   <chr>  <dbl>    <dbl>
#>  1 Mauna Loa CO2     arima   2021     414.
#>  2 Mauna Loa CO2     arima   2021     415.
#>  3 Mauna Loa CO2     arima   2021     417.
#>  4 Mauna Loa CO2     arima   2022     416.
#>  5 Mauna Loa CO2     arima   2022     418.
#>  6 Mauna Loa CO2     arima   2022     419.
#>  7 Mauna Loa CO2     arima   2023     418.
#>  8 Mauna Loa CO2     arima   2023     420.
#>  9 Mauna Loa CO2     arima   2023     421.
#> 10 Mauna Loa CO2     arima   2024     421.
#> # ... with 170 more rows
```

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

```
#> # A tibble: 2 x 5
#>   City      Measure .model lb_stat lb_pvalue
#>   <chr>     <fct>   <chr>    <dbl>     <dbl>
#> 1 Mauna Loa CO2     arima     25.5    0.702
#> 2 Mauna Loa CO2     ets       41.3    0.0814
```

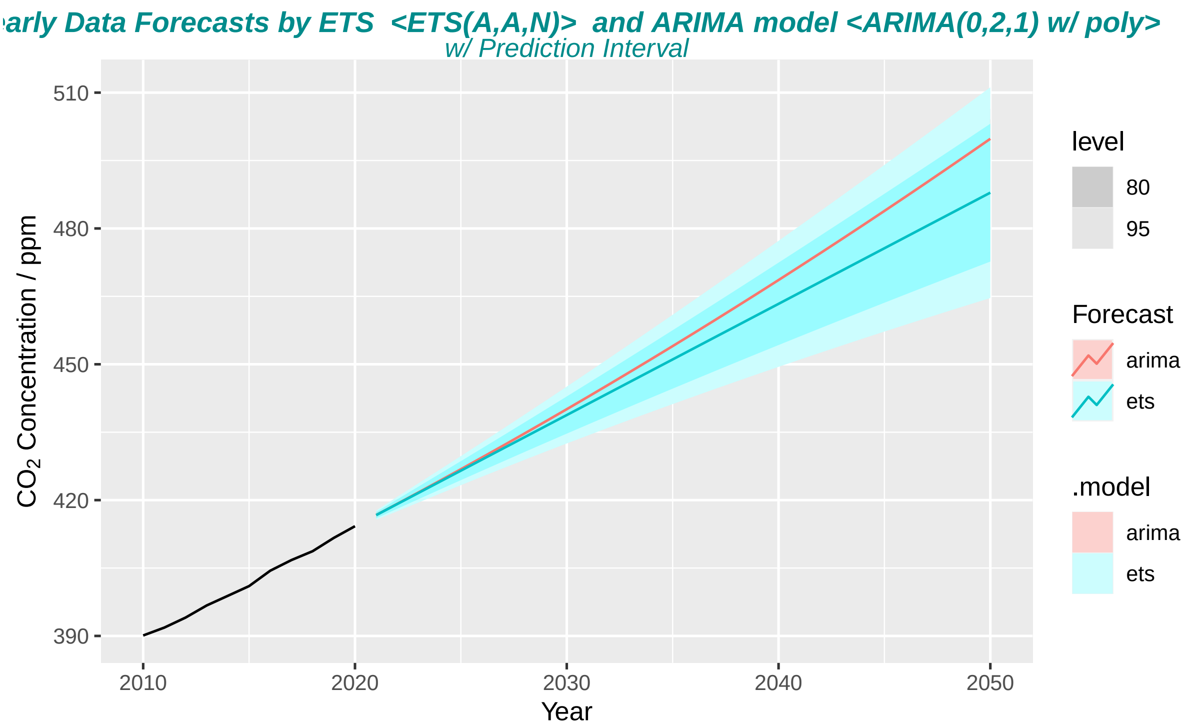
5 Yearly Data Forecasts with ARIMA and ETS

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.

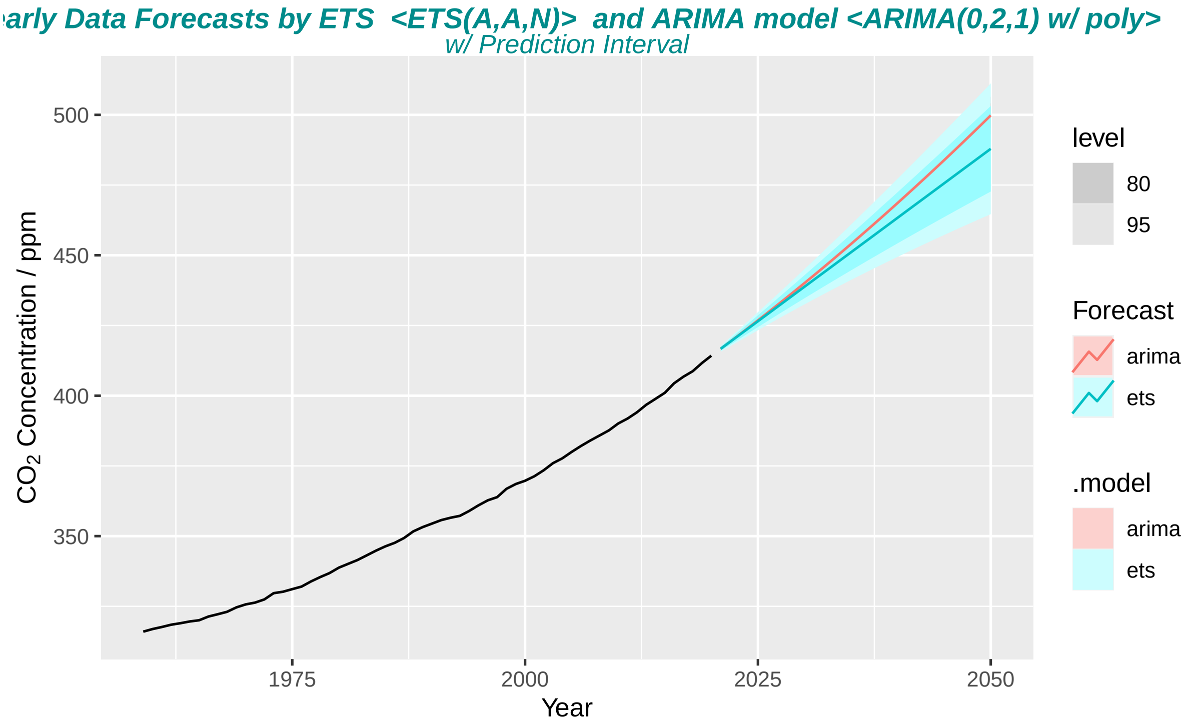
The ETS model $< ETS(A, A, N) >$ with seasonal term change “A” -> “N” is chosen. For ARIMA models the seasonal term (P,D,Q)m has to be taken out and an optimal ARIMA(p,1,q) with one differencing (d=1) is selected. However, for Mauna Loa two times differncing had to be selected $CO_2 <ARIMA(0,2,1) w/ poly>. For Temperature and Precipitation the same model as for monthly data can be taken by leaving out the seasonal term $< ARIMA(0, 1, 2)w/drift >$.

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




```
#> # A tsibble: 6 x 8 [1Y]
#> # Key:       City, Measure, .model [2]
#> # Groups:    City, Measure, .model [2]
#>   City  Measure .model  Year     Year_avg .mean                   `80%`
#>   <chr> <fct>   <chr>  <dbl>       <dist> <dbl>                  <hilo>
#> 1 Maun~ CO2     arima   2021 N(417, 0.24)  417. [416.0702, 417.3326]80
#> 2 Maun~ CO2     arima   2022 N(419, 0.49)  419. [418.2906, 420.0936]80
#> 3 Maun~ CO2     arima   2023 N(422, 0.76)  422. [420.5947, 422.8245]80
#> 4 Maun~ CO2     ets     2021 N(417, 0.28)  417. [416.0067, 417.3681]80
#> 5 Maun~ CO2     ets     2022 N(419, 0.66)  419. [418.1020, 420.1846]80
#> 6 Maun~ CO2     ets     2023  N(422, 1.2)  422. [420.2111, 422.9872]80
#> # ... with 1 more variable: `95%` <hilo>
#> # A tsibble: 6 x 8 [1Y]
#> # Key:       City, Measure, .model [2]
#> # Groups:    City, Measure, .model [2]
#>   City  Measure .model  Year    Year_avg .mean                   `80%`
#>   <chr> <fct>   <chr>  <dbl>      <dist> <dbl>                  <hilo>
#> 1 Maun~ CO2     arima   2048   N(493, 10)  493. [489.2074, 497.4967]80
#> 2 Maun~ CO2     arima   2049   N(497, 11)  497. [492.3207, 500.8113]80
#> 3 Maun~ CO2     arima   2050   N(500, 11)  500. [495.4612, 504.1522]80
#> 4 Maun~ CO2     ets     2048  N(483, 118)  483. [469.0817, 496.9106]80
#> 5 Maun~ CO2     ets     2049  N(485, 129)  485. [470.8901, 500.0140]80
#> 6 Maun~ CO2     ets     2050  N(488, 141)  488. [472.6888, 503.1270]80
#> # ... with 1 more variable: `95%` <hilo>
```



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

```
#> # A tibble: 2 x 5
#>   City      Measure .model lb_stat lb_pvalue
#>   <chr>     <fct>   <chr>    <dbl>     <dbl>
#> 1 Mauna Loa CO2     arima     54.8  0.00371
#> 2 Mauna Loa CO2     ets       64.0  0.000291
```

6 Backup