Climate Data Forecasting -

Atmospheric CO_2 Concentration / Temperature / Precipitation

Wolfgang Vollmer

2023-02-20

Contents

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

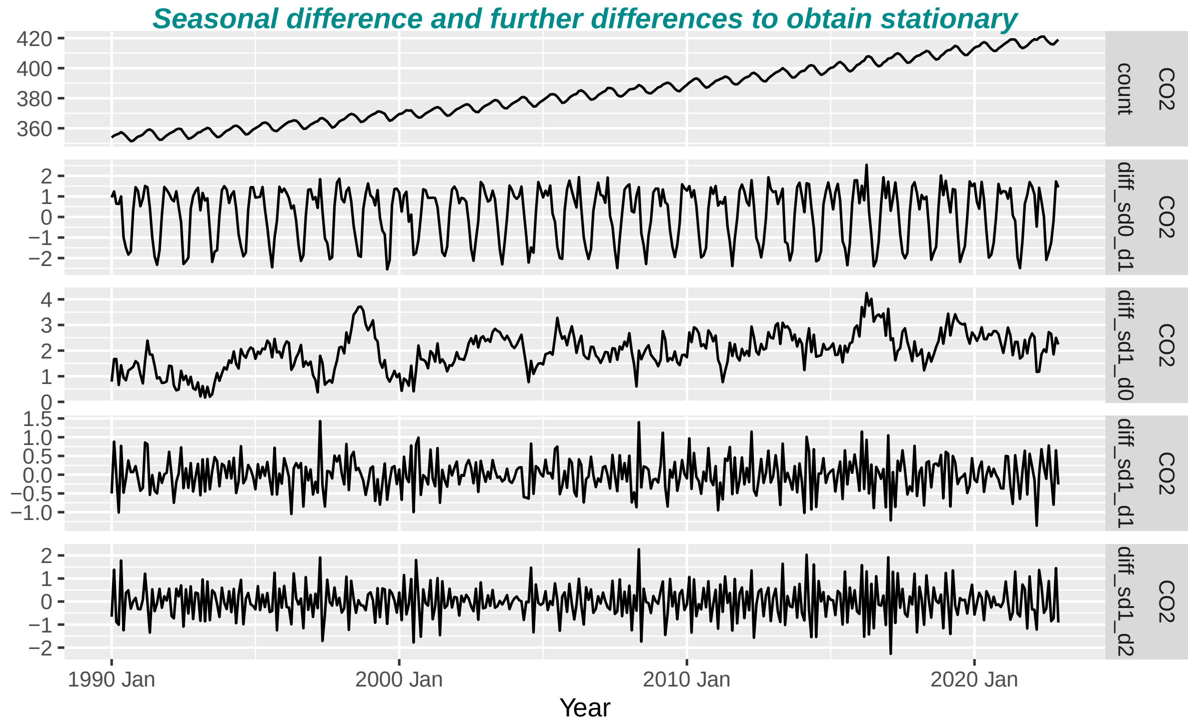
The model fit residuals have to be stationary. For good forecasting this has to be verified with residual diagnostics.

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       7109.         0
#> Ljung-Box test on (difference(count, 12))
#> # A tibble: 1 x 3
#>   Measure lb_stat lb_pvalue
#>   <fct>     <dbl>     <dbl>
#> 1 CO2       3030.         0
#> Ljung-Box test on (difference(count, 12) + difference())
#> # A tibble: 1 x 3
#>   Measure lb_stat lb_pvalue
#>   <fct>     <dbl>     <dbl>
#> 1 CO2        79.2  7.21e-13
```

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

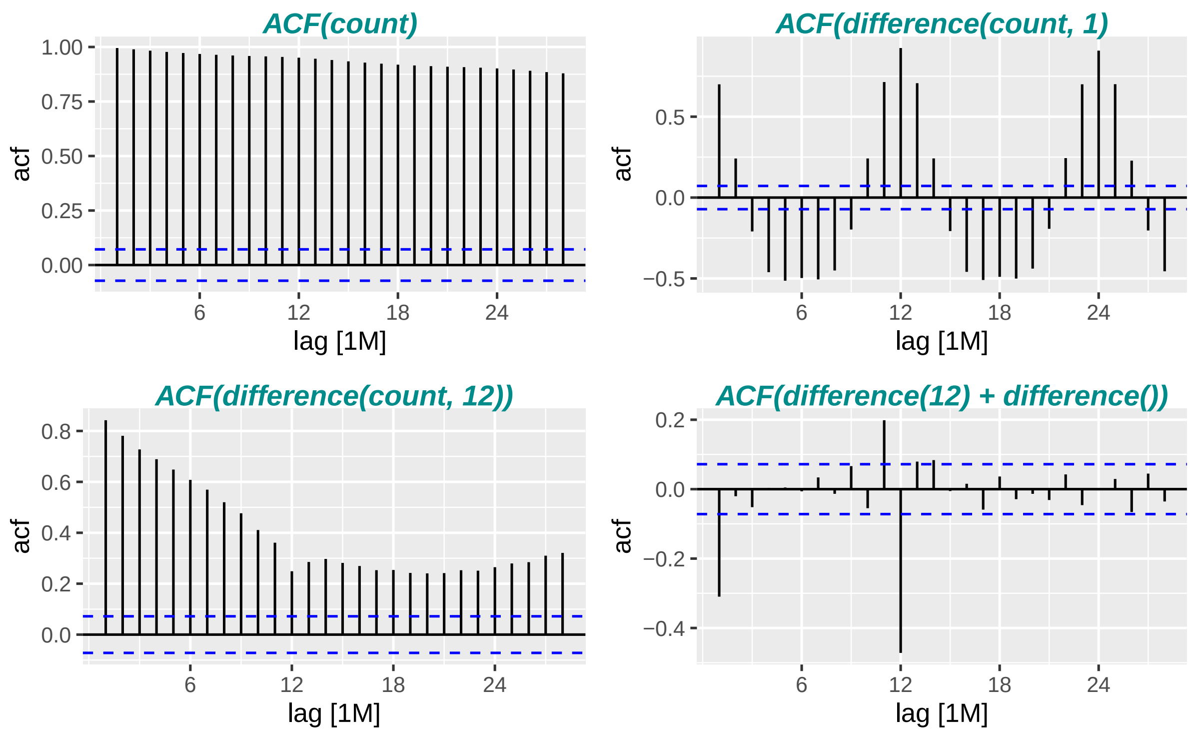
kpss test of stationary
Null Hypothesis of stationary in a given time series
=> H_0 to be rejected for $p < \alpha = 0.05$

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.

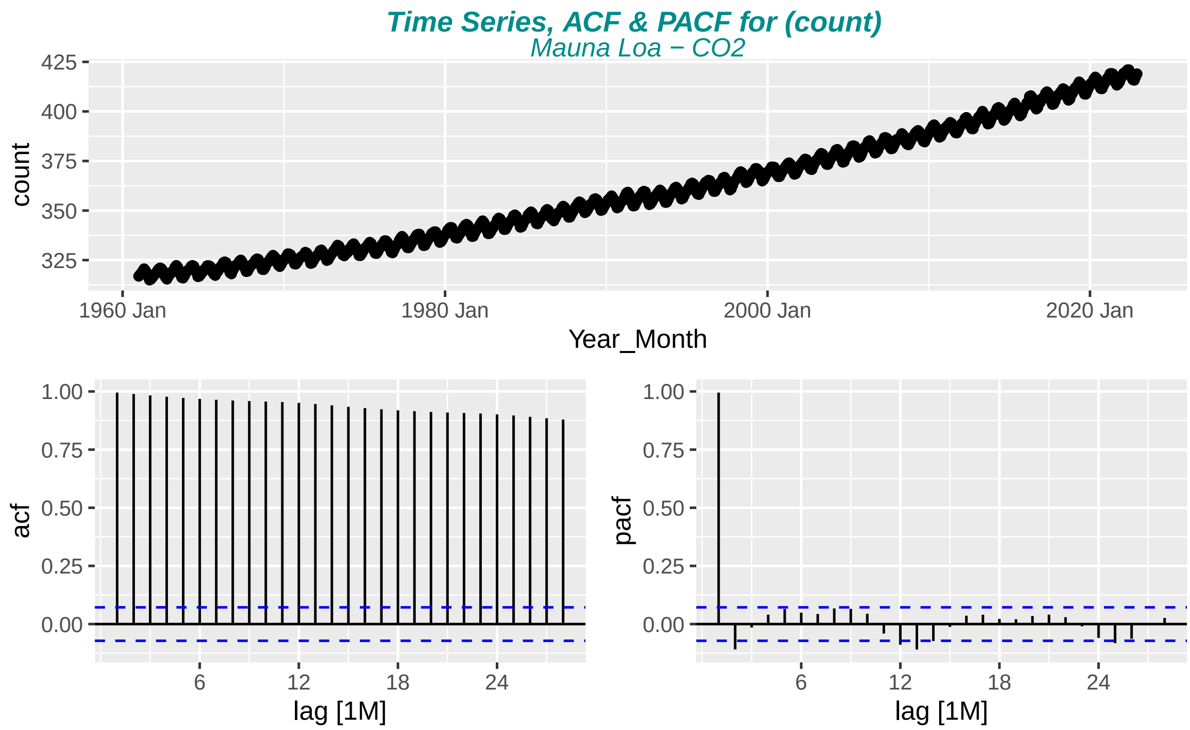
- returns 1 => for stationarity one seasonal difference rsp. difference is required

```
#> 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          5.14        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       0.00724         0.1       0      0
```

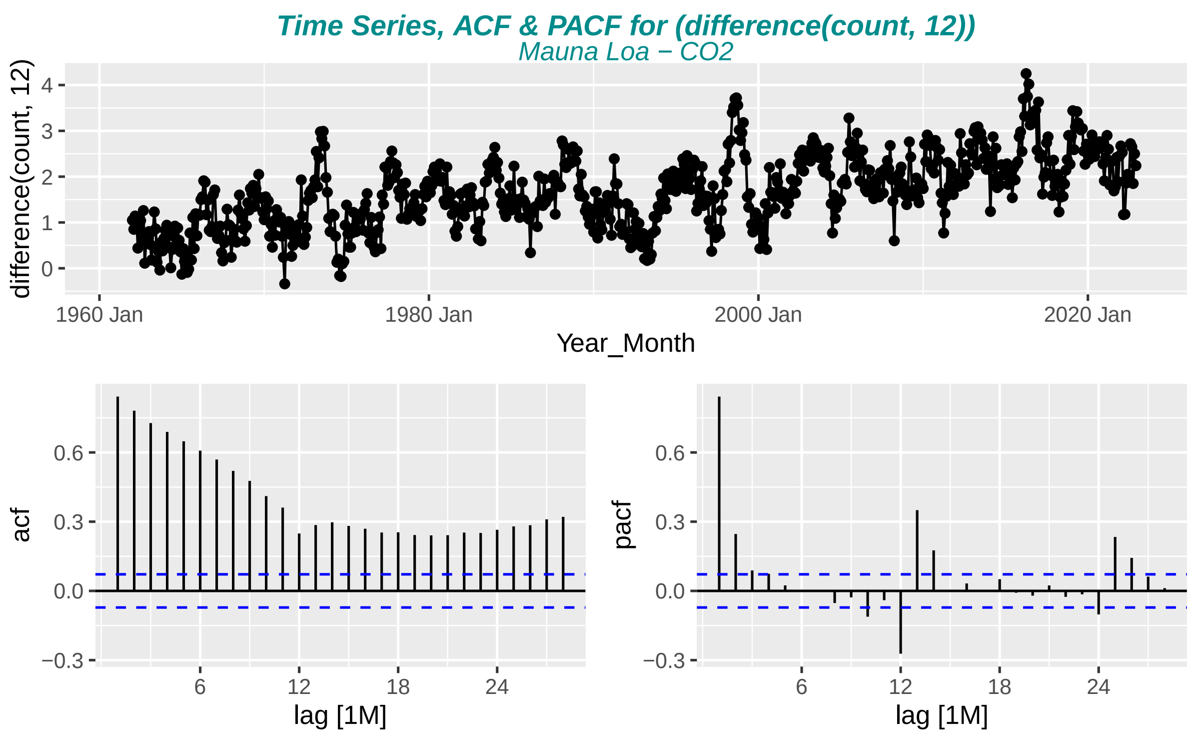
1.1.3 ACF Plots of Differences



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

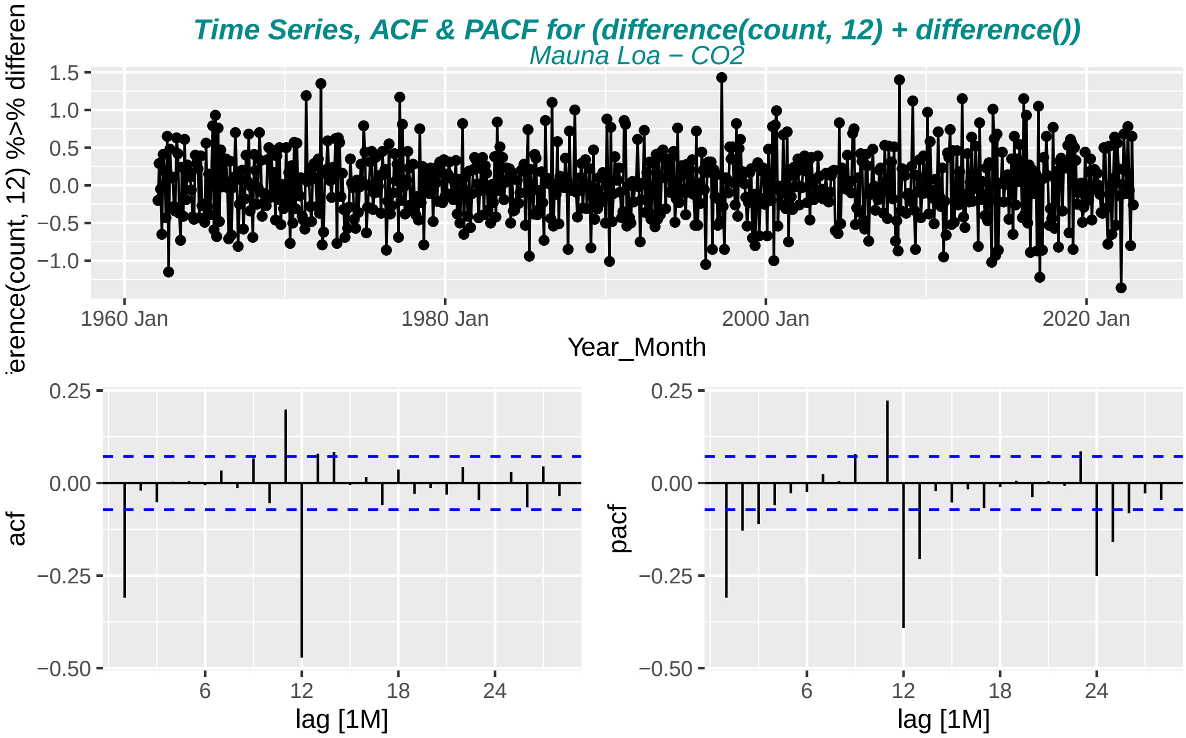



```
#> # A tibble: 1 x 2
#>     Sum  Mean
#>   <dbl> <dbl>
#> 1 1211.  1.65
```




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

```
#>    <dbl> <dbl>
#> 1 1211.  1.65
```




```
#> # A tibble: 1 x 2
#>      Sum    Mean
#>    <dbl>   <dbl>
#> 1   1.19 0.00163
```

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").

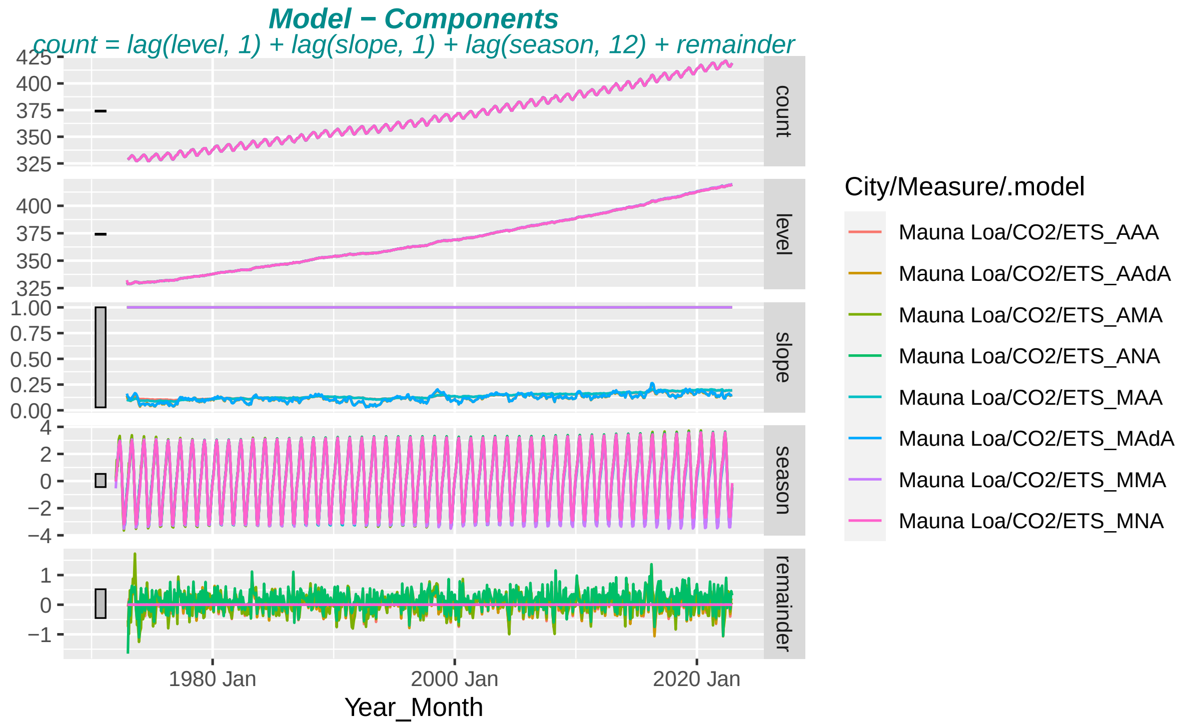
For CO2 and Temperature Data we have a clear seasonal pattern and seasonal term adds always a (more or less) fix amount on level and trend component. Therefore "A" additve term is chosen. For Precipitation the seasonal patttern is only slight. Indead, a multiplicative seasonal term results in "exploding" 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 Loa CO2     ETS(coun~  0.103  -1228. 2491. 2492. 2565. 0.100 0.141 0.247
#> Series: count
#> Model: ETS(A,A,A)
#>   Smoothing parameters:
#>     alpha = 0.5752064
#>     beta  = 0.006657306
#>     gamma = 0.08427619
#>
#>   Initial states:
#>      l[0]      b[0]       s[0]    s[-1]      s[-2]     s[-3]     s[-4]     s[-5]
#>  328.5453 0.1129828 -0.8389104 -1.94534 -3.099749 -3.065103 -1.379817 0.6818637
#>     s[-6]    s[-7]    s[-8]    s[-9]    s[-10]    s[-11]
#>  2.193653 2.899279 2.485594 1.475993 0.6359707 -0.043435
#>
#>   sigma^2:  0.1027
#>
#>      AIC     AICc      BIC
#> 2490.599 2491.651 2565.347
#> 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 Loa CO2     ETS_A~ 1.03e-1  -1228. 2491. 2492. 2565. 0.100 0.141 2.47e-1
#> 2 Mauna Loa CO2     ETS_M~ 7.60e-7  -1228. 2491. 2492. 2566. 0.102 0.147 6.74e-4
#> 3 Mauna Loa CO2     ETS_M~ 8.08e-7  -1246. 2528. 2529. 2607. 0.108 0.161 6.99e-4
#> 4 Mauna Loa CO2     ETS_A~ 1.11e-1  -1250. 2536. 2537. 2615. 0.107 0.163 2.57e-1
#> 5 Mauna Loa CO2     ETS_M~ 9.08e-7  -1282. 2597. 2598. 2672. 0.118 0.162 7.35e-4
#> 6 Mauna Loa CO2     ETS_A~ 1.32e-1  -1302. 2639. 2640. 2714. 0.128 0.166 2.78e-1
#> 7 Mauna Loa CO2     ETS_A~ 1.50e-1  -1342. 2714. 2715. 2780. 0.146 0.285 3.05e-1
#> 8 Mauna Loa CO2     ETS_M~ 1.26e-6  -1380. 2791. 2792. 2857. 0.164 0.303 8.36e-4
```



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

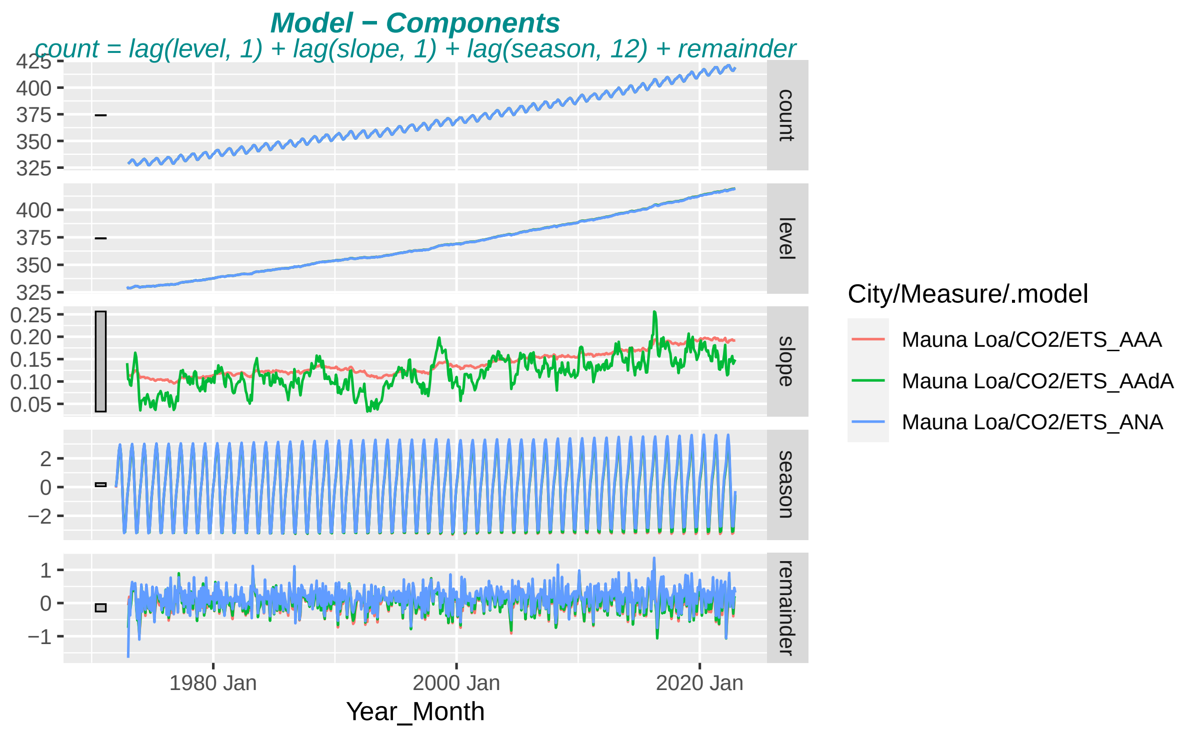
- MAE (Mean absolute error, will lead to forecasts of the median) and
- RMSE (Root mean squared error, lead to forecasts of the mean)

```
#> # A tibble: 8 x 12
#>   City      Measure .model  .type      ME  RMSE   MAE     MPE   MAPE  MASE RMSSE
#>   <chr>     <fct>   <chr>   <chr>   <dbl> <dbl> <dbl>   <dbl>  <dbl> <dbl> <dbl>
#> 1 Mauna Loa CO2     ETS_AAA Trai~ 0.0197  0.316 0.247 5.09e-3 0.0670 0.136 0.162
#> 2 Mauna Loa CO2     ETS_MAA Trai~ 0.0168  0.319 0.249 4.42e-3 0.0674 0.137 0.163
#> 3 Mauna Loa CO2     ETS_AA~ Trai~ 0.0536  0.328 0.257 1.43e-2 0.0697 0.142 0.168
#> 4 Mauna Loa CO2     ETS_MA~ Trai~ 0.0541  0.328 0.258 1.44e-2 0.0699 0.142 0.168
#> 5 Mauna Loa CO2     ETS_MMA Trai~ 0.00190 0.344 0.270 4.66e-4 0.0735 0.149 0.176
#> 6 Mauna Loa CO2     ETS_AMA Trai~ 0.0261  0.358 0.278 6.67e-3 0.0759 0.153 0.183
#> 7 Mauna Loa CO2     ETS_ANA Trai~ 0.170   0.382 0.305 4.54e-2 0.0826 0.168 0.196
#> 8 Mauna Loa CO2     ETS_MNA Trai~ 0.157   0.405 0.308 4.19e-2 0.0836 0.170 0.207
#> # ... with 1 more variable: ACF1 <dbl>
```

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

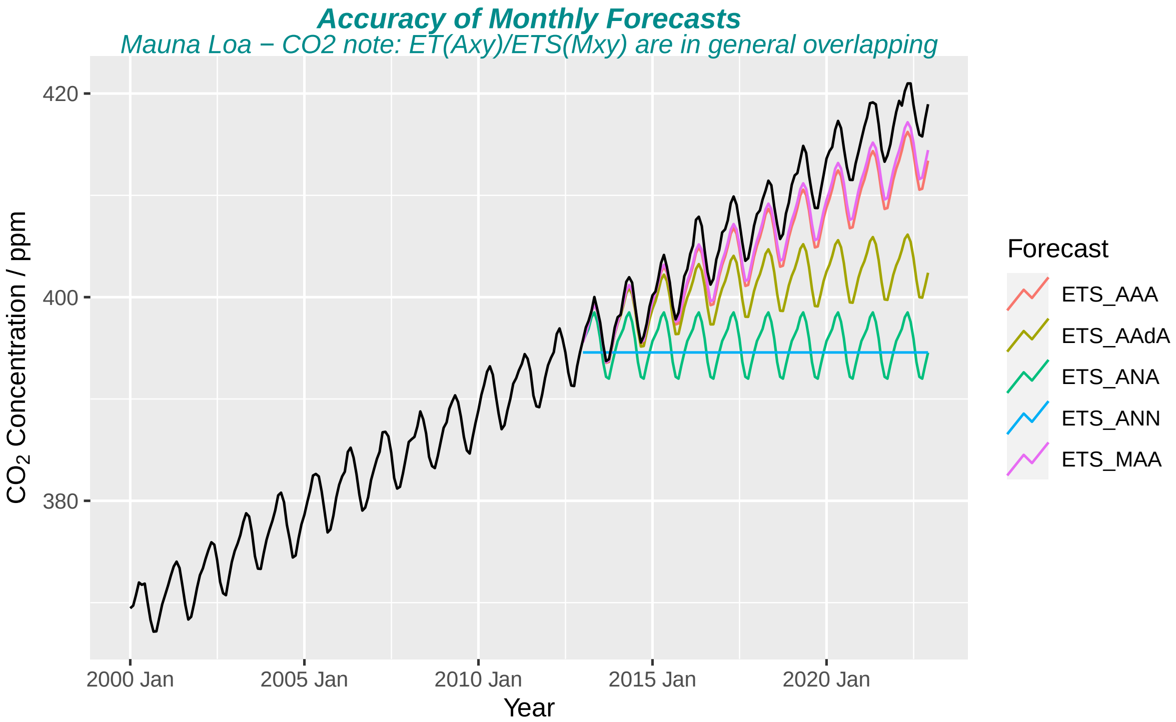
```
#> Null Hypothesis of independence/white noise for residuals - for p < 0.05: reject H_0
#> # A tibble: 8 x 5
#>   City      Measure .model    lb_stat    lb_pvalue
#>   <chr>     <fct>   <chr>       <dbl>        <dbl>
#> 1 Mauna Loa CO2     ETS_AAA      45.9 0.0318
#> 2 Mauna Loa CO2     ETS_AAdA     58.9 0.00125
#> 3 Mauna Loa CO2     ETS_MAdA     59.7 0.000997
#> 4 Mauna Loa CO2     ETS_ANA      60.2 0.000877
#> 5 Mauna Loa CO2     ETS_MNA      60.3 0.000846
#> 6 Mauna Loa CO2     ETS_MAA      62.6 0.000444
#> 7 Mauna Loa CO2     ETS_MMA      83.9 0.000000543
#> 8 Mauna Loa CO2     ETS_AMA      93.8 0.0000000173
```

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

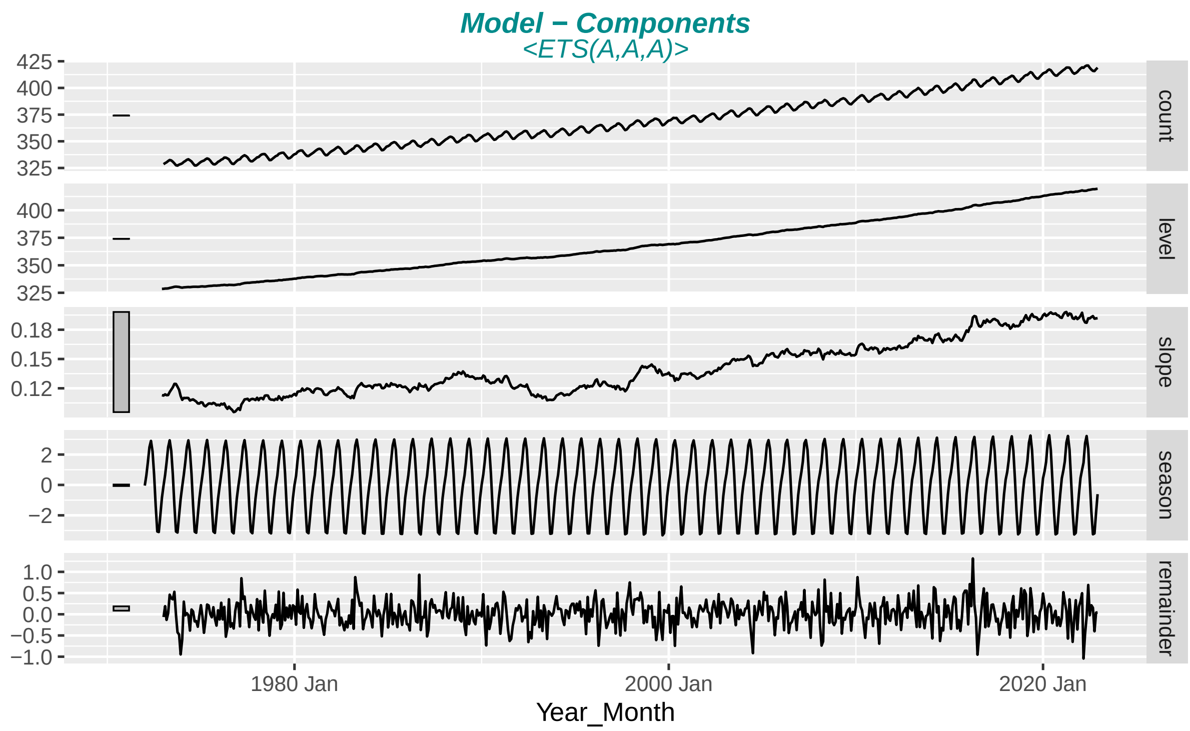
```
#> # A tibble: 5 x 12
#>   .model   City    Measure .type    ME  RMSE   MAE   MPE  MAPE  MASE RMSSE  ACF1
#>   <chr>    <chr>   <fct>   <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
#> 1 ETS_MAA  Mauna ~ CO2     Test   2.32  2.74  2.32 0.562 0.564  1.41  1.54 0.941
#> 2 ETS_AAA  Mauna ~ CO2     Test   2.83  3.34  2.83 0.687 0.687  1.71  1.87 0.953
#> 3 ETS_AAdA Mauna ~ CO2     Test   6.94  8.65  6.94 1.68  1.68   4.20  4.86 0.975
#> 4 ETS_ANA  Mauna ~ CO2     Test  12.4  14.3  12.4  3.00  3.00   7.48  8.03 0.974
#> 5 ETS_ANN  Mauna ~ CO2     Test  13.2  15.1  13.2  3.20  3.20   7.98  8.48 0.962
```

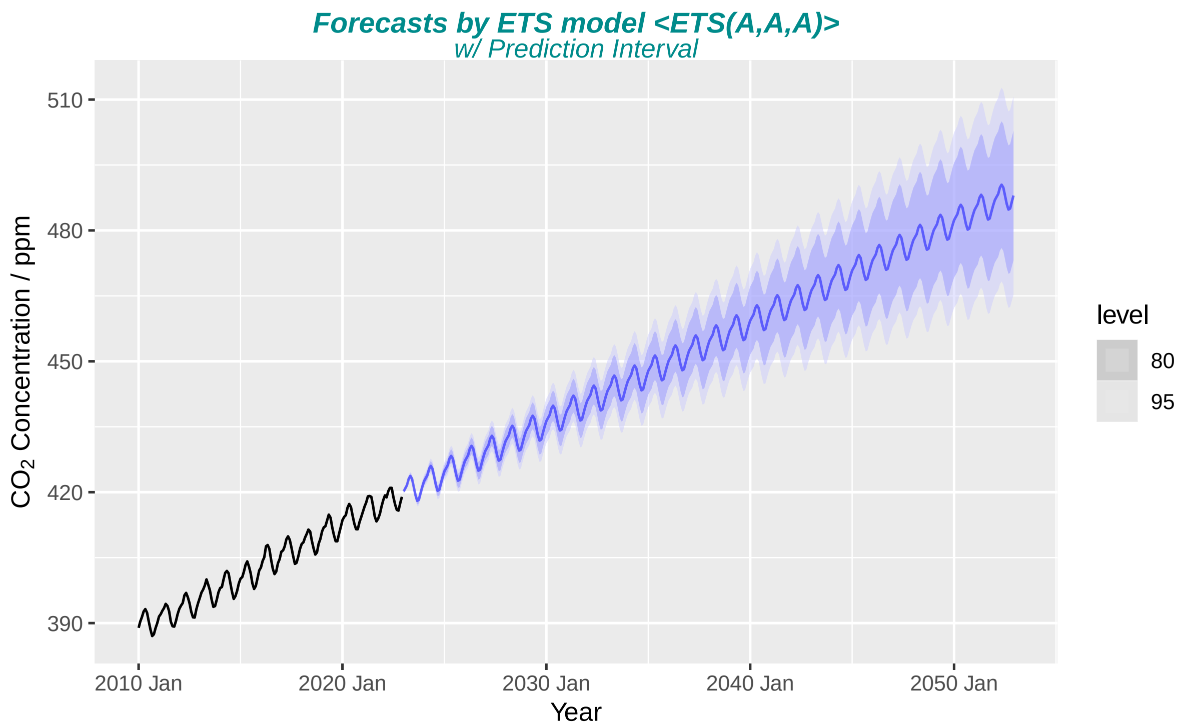


2.2 Forecasting with selected ETS model <ETS(A,A,A)>

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 = 0.5752064
#>     beta  = 0.006657306
#>     gamma = 0.08427619
#>
#>   Initial states:
#>      l[0]      b[0]       s[0]     s[-1]      s[-2]      s[-3]      s[-4]      s[-5]
#>  328.5453 0.1129828 -0.8389104 -1.94534 -3.099749 -3.065103 -1.379817 0.6818637
#>     s[-6]    s[-7]    s[-8]    s[-9]    s[-10]     s[-11]
#>  2.193653 2.899279 2.485594 1.475993 0.6359707 -0.043435
#>
#>   sigma^2:  0.1027
#>
#>      AIC     AICc      BIC
#> 2490.599 2491.651 2565.347
```

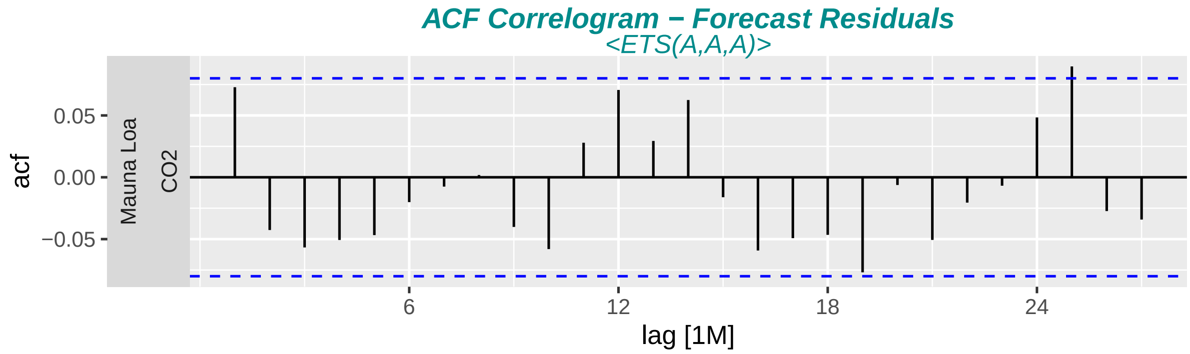


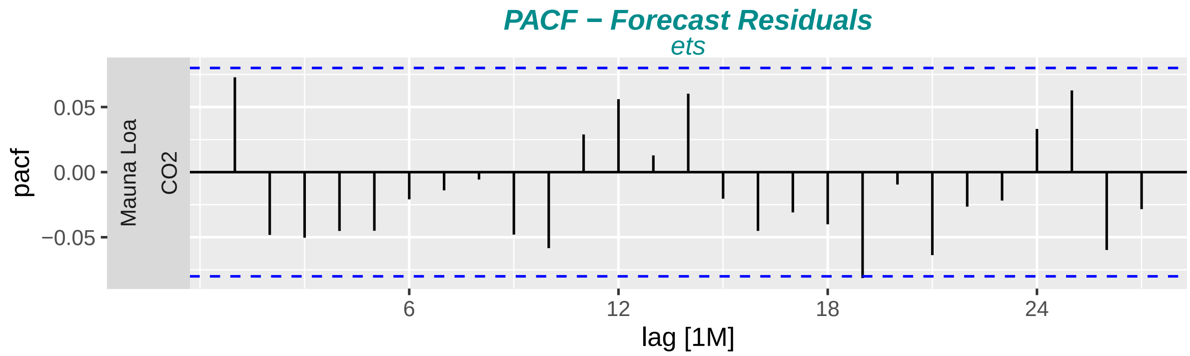


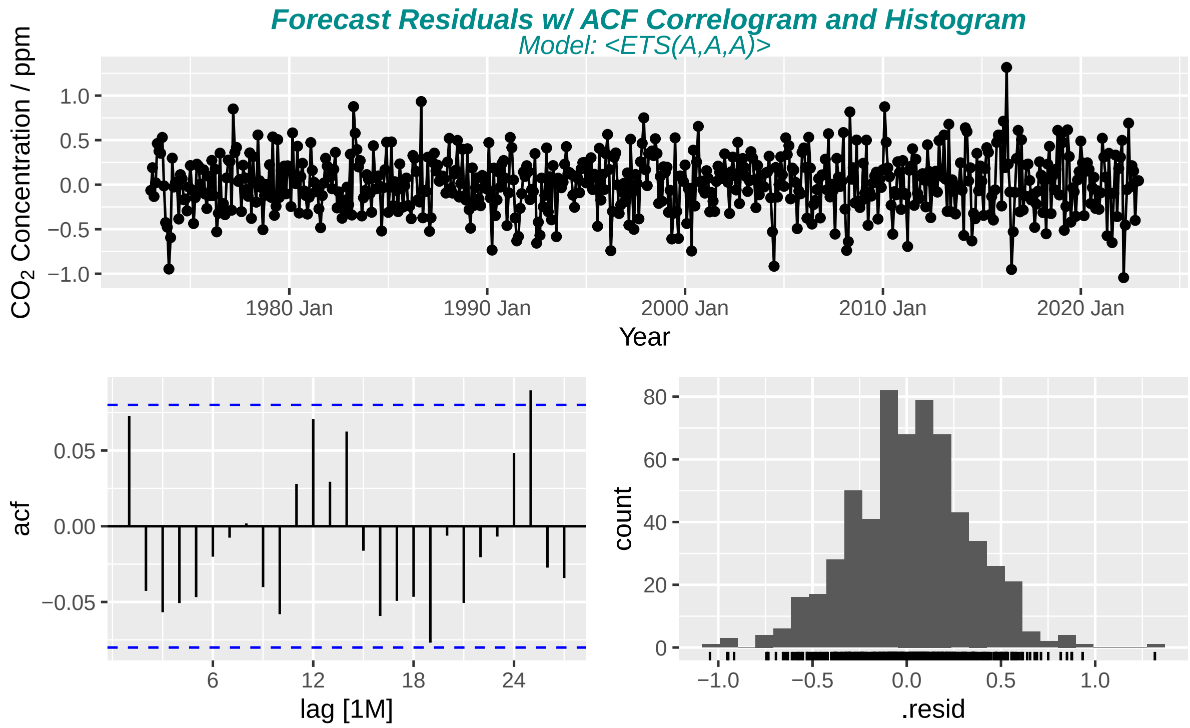
2.2.2 Residual Stationarity

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

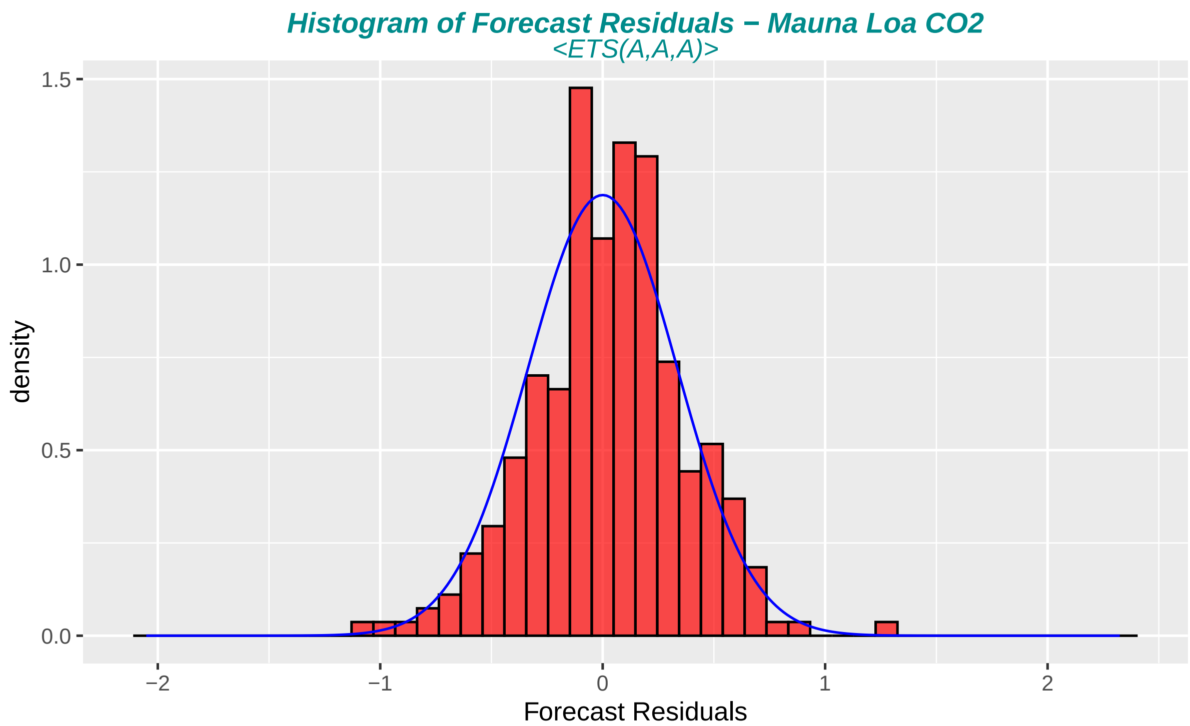






2.2.3 Histogram of forecast residuals with overlaid normal curve

```
#> 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       44.0    0.0473
```



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

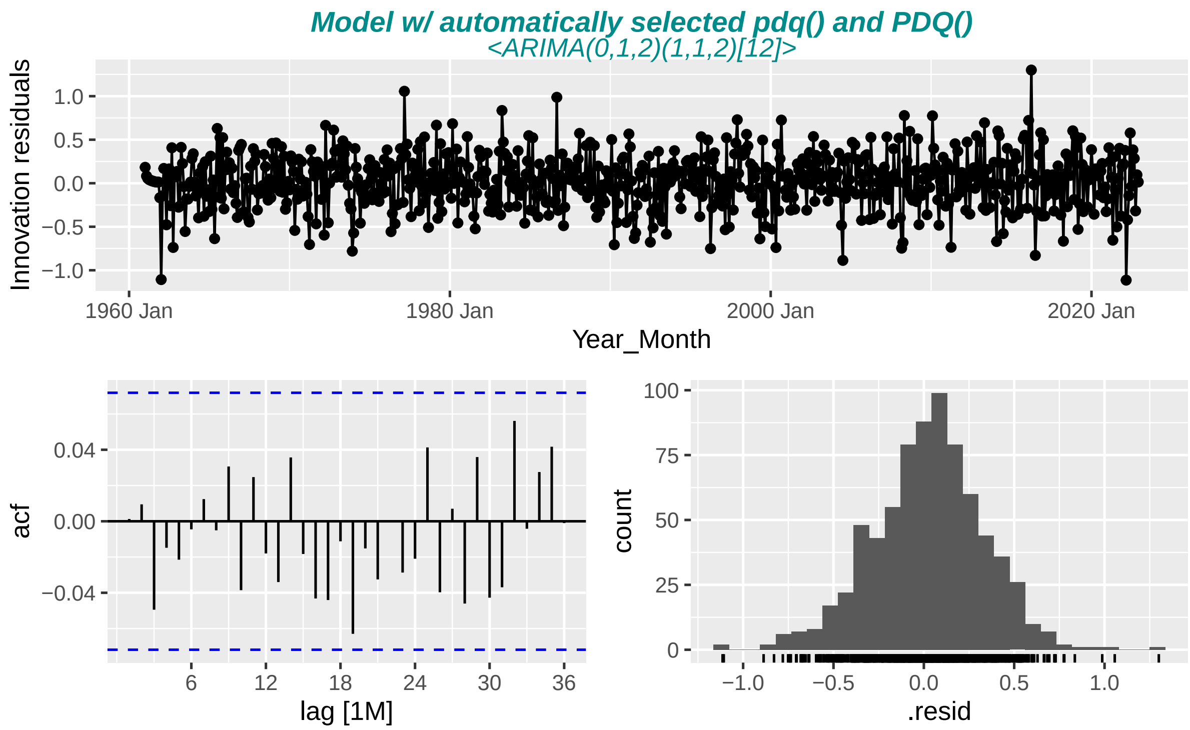
Non-seasonal ARIMA models are generally denoted ARIMA(p,d,q) where parameters p, d, and q are non-negative integers, * p is the order (number of time lags) of the autoregressive model * d is the degree of differencing (number of times the data have had past values subtracted) * q is the order of the 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  0.0985   -189.  391.  391.  418. <cpl [12]> <cpl>
#> Series: count
#> Model: ARIMA(0,1,2)(1,1,2)[12]
```

```
#>
#> Coefficients:
#>           ma1      ma2     sar1     sma1     sma2
#>       -0.3710  -0.0618  -0.5001  -0.3583  -0.4427
#> s.e.   0.0368   0.0361      NaN      NaN      NaN
#>
#> sigma^2 estimated as 0.09849:  log likelihood=-189.28
#> AIC=390.55   AICc=390.67   BIC=418.12
```




```
#> 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_ro~1 ma_ro~2
#>    <chr>     <fct>   <chr>       <dbl>   <dbl> <dbl> <dbl> <dbl> <list>  <list>
#>  1 Mauna Loa CO2     ARIMA_111~  0.101   -160.  331.  331.  353. <cpl>   <cpl>
#>  2 Mauna Loa CO2     ARIMA_012~  0.101   -161.  331.  331.  353. <cpl>   <cpl>
#>  3 Mauna Loa CO2     ARIMA_211~  0.101   -161.  332.  332.  353. <cpl>   <cpl>
#>  4 Mauna Loa CO2     ARIMA_111~  0.101   -160.  333.  333.  359. <cpl>   <cpl>
#>  5 Mauna Loa CO2     ARIMA_210~  0.130   -230.  468.  468.  485. <cpl>   <cpl>
#>  6 Mauna Loa CO2     ARIMA_100~  0.130   -230.  469.  469.  491. <cpl>   <cpl>
#>  7 Mauna Loa CO2     ARIMA_200~  0.130   -230.  469.  469.  491. <cpl>   <cpl>
#>  8 Mauna Loa CO2     ARIMA_100~  0.133   -237.  484.  485.  506. <cpl>   <cpl>
#>  9 Mauna Loa CO2     ARIMA_010~  0.148   -268.  539.  539.  548. <cpl>   <cpl>
#> 10 Mauna Loa CO2     ARIMA_012~  0.167   -303.  612.  612.  625. <cpl>   <cpl>
#> 11 Mauna Loa CO2     ARIMA_111~  0.167   -303.  612.  612.  625. <cpl>   <cpl>
#> 12 Mauna Loa CO2     ARIMA_110~  0.173   -313.  630.  630.  639. <cpl>   <cpl>
#> # ... with abbreviated variable names 1: ar_roots, 2: ma_roots
```

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)

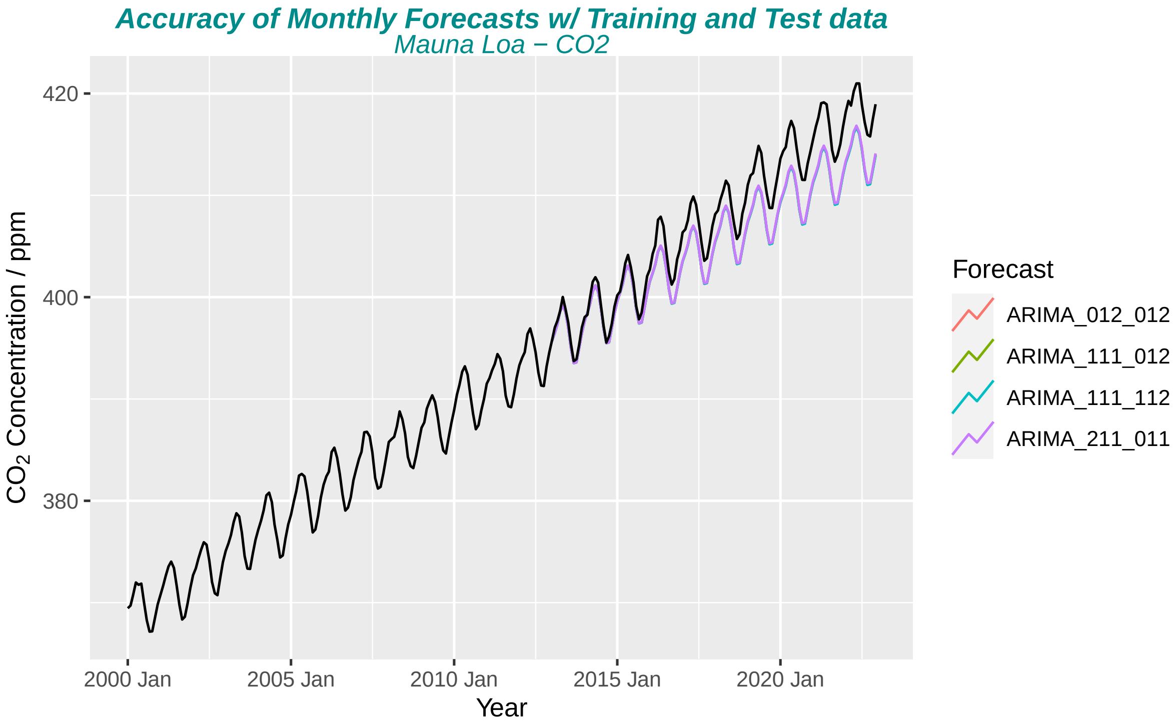
```
#> # A tibble: 14 x 12
#>    City      Measure .model   .type        ME    RMSE     MAE       MPE     MAPE
#>    <chr>     <fct>   <chr>    <chr>     <dbl>   <dbl>   <dbl>     <dbl>    <dbl>
#>  1 Mauna Loa CO2     ARIMA_1~ Trai~   0.0257    0.314   0.243   6.85e-3   0.0661
#>  2 Mauna Loa CO2     ARIMA_1~ Trai~   0.0257    0.314   0.243   6.86e-3   0.0660
#>  3 Mauna Loa CO2     ARIMA_0~ Trai~   0.0252    0.314   0.244   6.71e-3   0.0661
#>  4 Mauna Loa CO2     ARIMA_2~ Trai~   0.0259    0.314   0.244   6.91e-3   0.0661
#>  5 Mauna Loa CO2     ARIMA_1~ Trai~   0.00879   0.356   0.284   1.78e-3   0.0769
#>  6 Mauna Loa CO2     ARIMA_2~ Trai~   0.00879   0.356   0.284   1.78e-3   0.0769
#>  7 Mauna Loa CO2     ARIMA_2~ Trai~   0.00381   0.356   0.280   1.03e-3   0.0758
#>  8 Mauna Loa CO2     ARIMA_1~ Trai~   0.00725   0.360   0.283   1.11e-3   0.0769
#>  9 Mauna Loa CO2     ARIMA_0~ Trai~   0.00224   0.380   0.293   5.88e-4   0.0791
#> 10 Mauna Loa CO2     ARIMA_0~ Trai~   0.00290   0.404   0.312   7.76e-4   0.0846
#> 11 Mauna Loa CO2     ARIMA_1~ Trai~   0.00294   0.404   0.312   7.87e-4   0.0846
#> 12 Mauna Loa CO2     ARIMA_1~ Trai~   0.00191   0.411   0.322   4.94e-4   0.0871
#> 13 Mauna Loa CO2     ARIMA_3~ Trai~ NaN       NaN     NaN     NaN       NaN
#> 14 Mauna Loa CO2     ARIMA_0~ Trai~ NaN       NaN     NaN     NaN       NaN
#> # ... with 3 more variables: MASE <dbl>, RMSSE <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_112    23.3  8.01e- 1
#>  2 Mauna Loa CO2     ARIMA_111_012    23.4  8.00e- 1
#>  3 Mauna Loa CO2     ARIMA_012_012    23.4  7.97e- 1
#>  4 Mauna Loa CO2     ARIMA_211_011    23.7  7.85e- 1
#>  5 Mauna Loa CO2     ARIMA_100_110    81.0  1.43e- 6
#>  6 Mauna Loa CO2     ARIMA_200_110    81.0  1.43e- 6
#>  7 Mauna Loa CO2     ARIMA_210_110    86.2  2.41e- 7
#>  8 Mauna Loa CO2     ARIMA_100_210   113.   1.68e-11
#>  9 Mauna Loa CO2     ARIMA_010_110   152.   0
#> 10 Mauna Loa CO2     ARIMA_012_010   160.   0
#> 11 Mauna Loa CO2     ARIMA_110_010   189.   0
#> 12 Mauna Loa CO2     ARIMA_111_010   161.   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

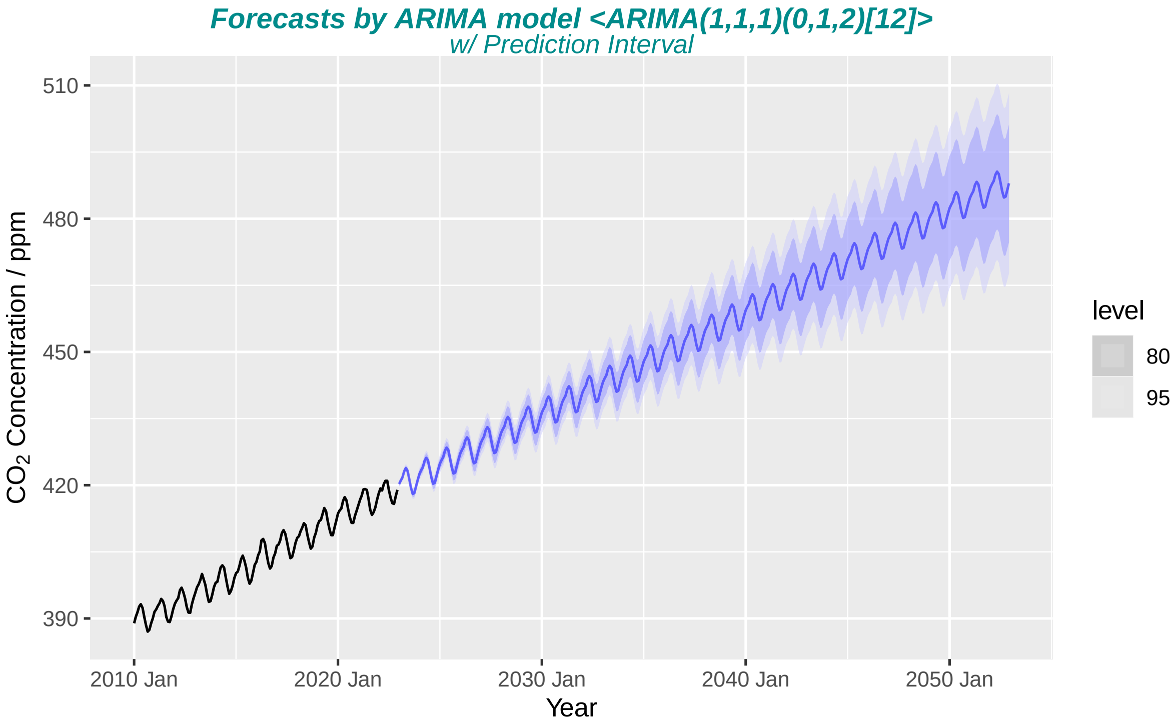
```
#> # A tibble: 4 x 12
#>   .model     City  Measure .type    ME  RMSE   MAE   MPE  MAPE  MASE RMSSE  ACF1
#>   <chr>      <chr> <fct>   <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
#> 1 ARIMA_211~ Maun~ CO2     Test   2.54  2.98  2.54 0.615 0.616  1.54  1.67 0.948
#> 2 ARIMA_012~ Maun~ CO2     Test   2.59  3.05  2.59 0.628 0.629  1.57  1.71 0.950
#> 3 ARIMA_111~ Maun~ CO2     Test   2.60  3.06  2.60 0.631 0.631  1.57  1.72 0.950
#> 4 ARIMA_111~ Maun~ CO2     Test   2.61  3.07  2.61 0.633 0.634  1.58  1.72 0.950
```



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

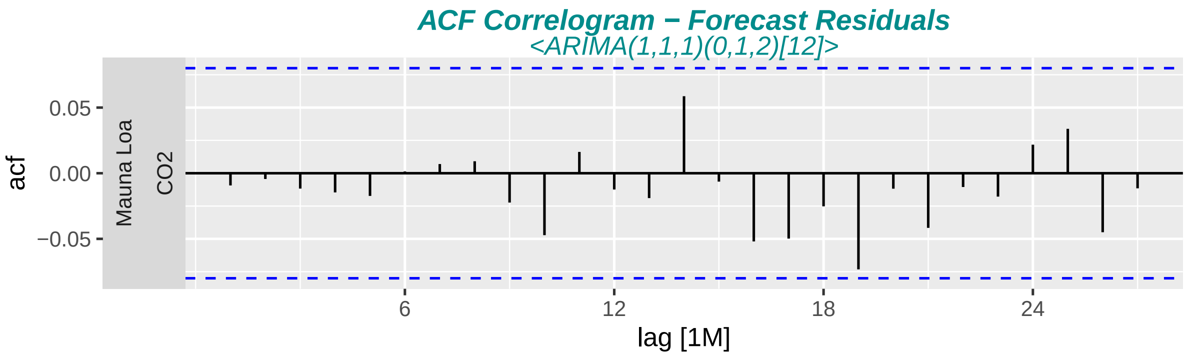
```
#> Provide model coefficients by report(fit_model)
#> Series: count
#> Model: ARIMA(1,1,1)(0,1,2)[12]
#>
#> Coefficients:
#>          ar1      ma1     sma1     sma2
#>       0.1824  -0.5440  -0.8378  -0.0360
#> s.e.  0.1024   0.0875   0.0423   0.0404
#>
#> sigma^2 estimated as 0.1013:  log likelihood=-160.43
#> AIC=330.87   AICc=330.97   BIC=352.74
```

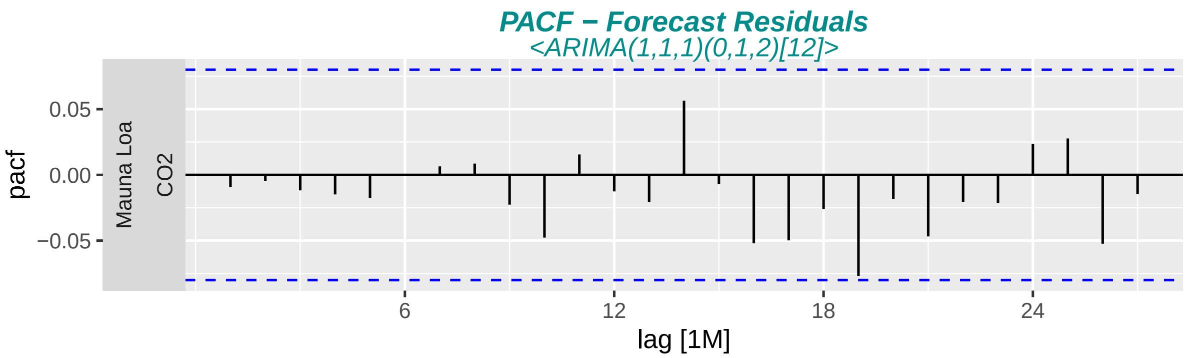


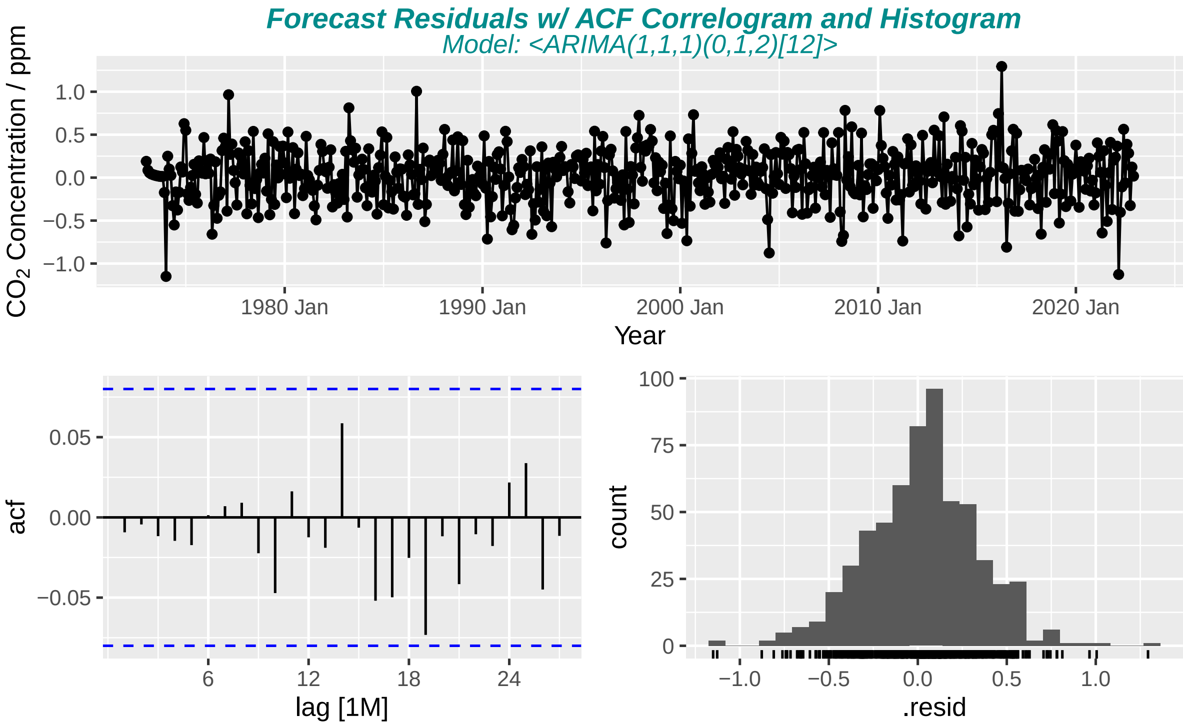
3.2.2 Residual Stationarity

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

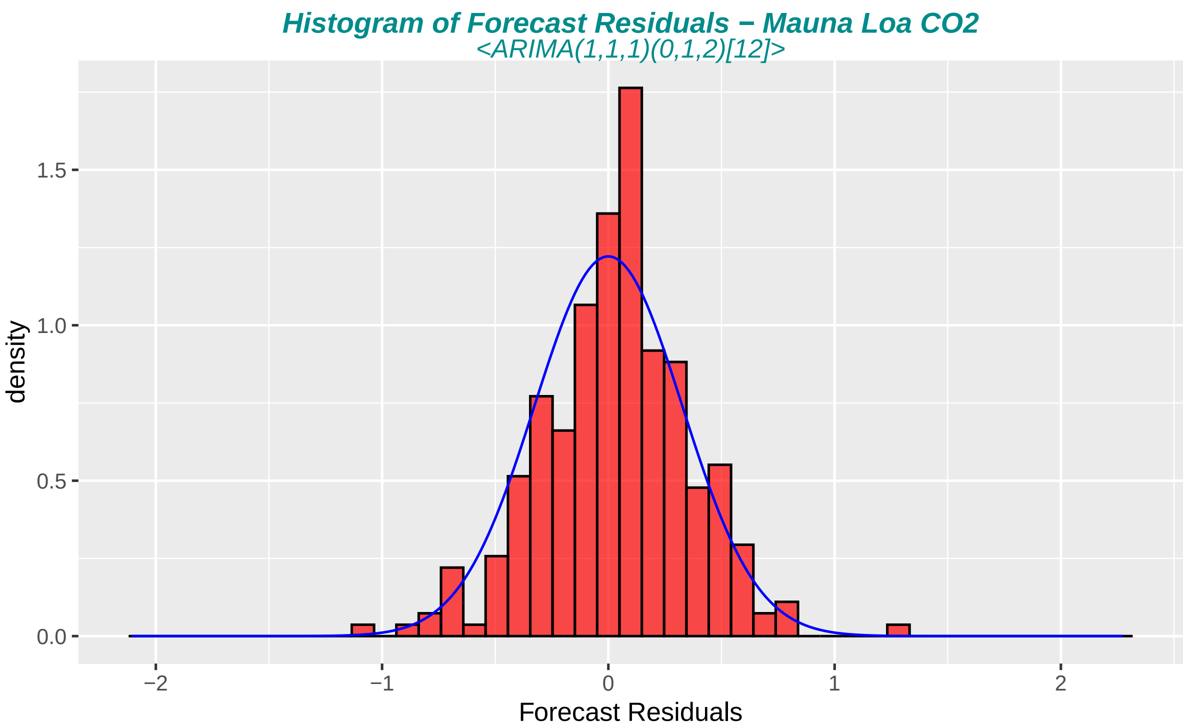






3.2.3 Histogram of forecast residuals with overlaid normal curve

```
#> 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     26.1     0.669
```



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.

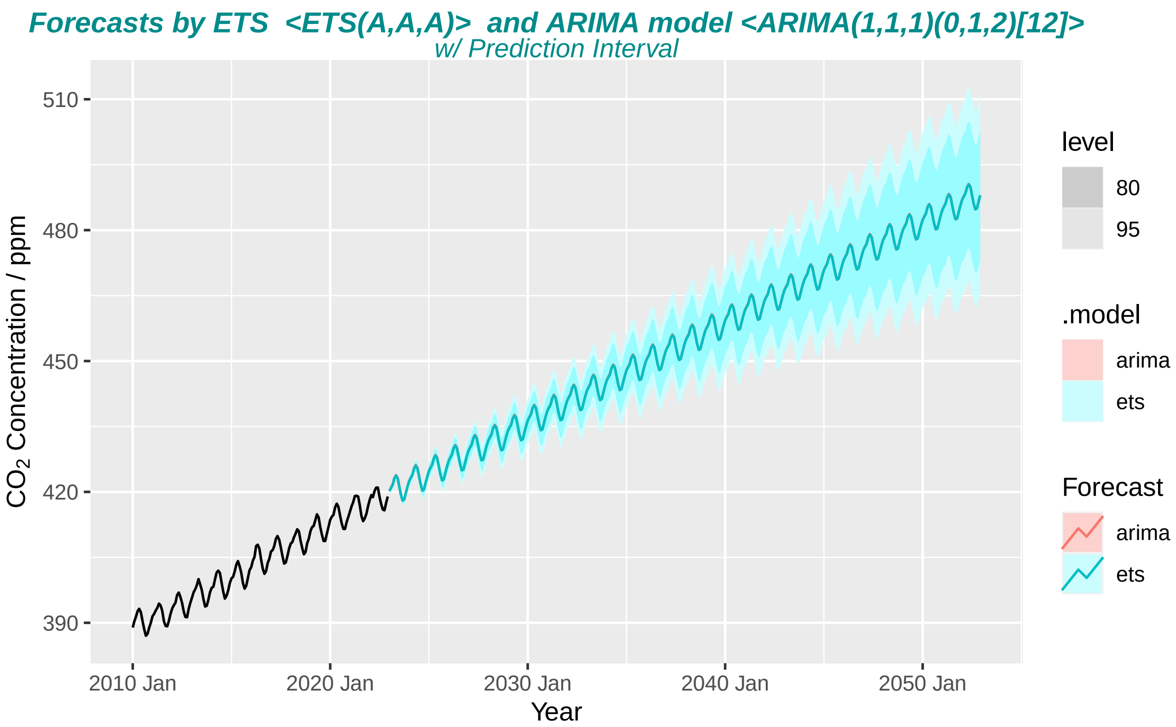
- Residual Accuracy with one-step-ahead fitted residuals
- Forecast Accuracy with Training/Test Data

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

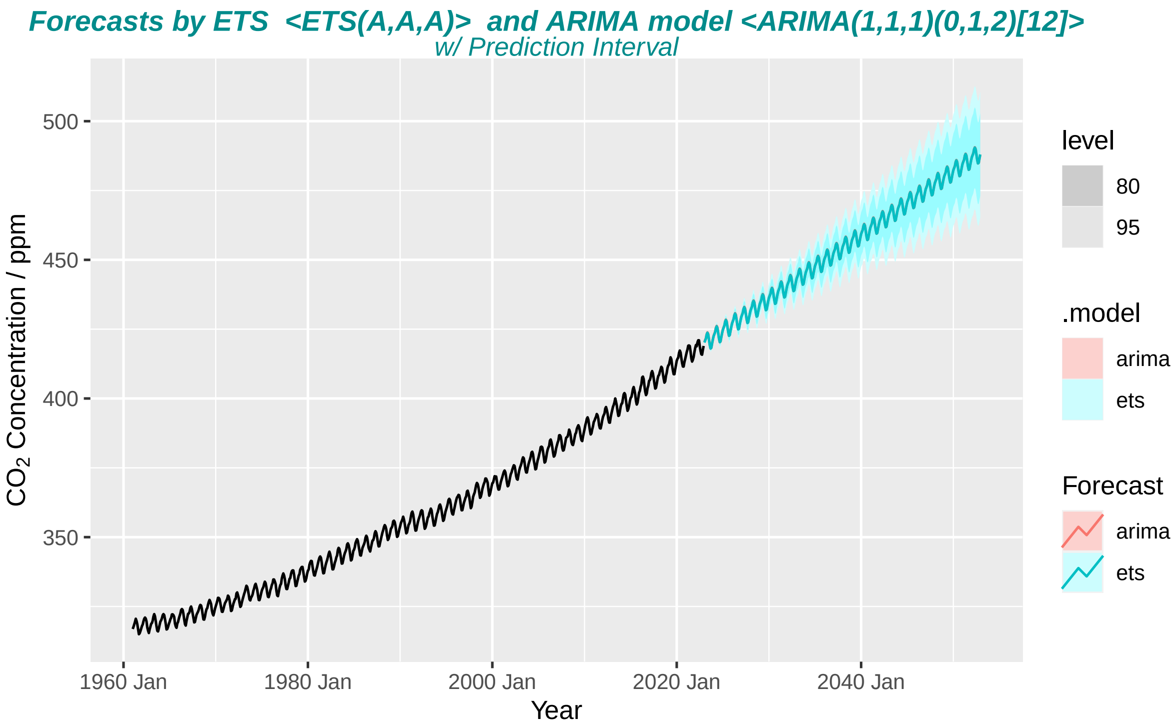
```
#> # A tibble: 4 x 12
#>   City      Measure .model   .type     ME  RMSE   MAE     MPE   MAPE  MASE RMSSE
#>   <chr>     <fct>   <chr>    <chr>  <dbl> <dbl> <dbl>   <dbl>  <dbl> <dbl> <dbl>
#> 1 Mauna Loa CO2     ets      Trai~ 0.0197 0.316 0.247 0.00509 0.0670 0.136 0.162
#> 2 Mauna Loa CO2     arima    Trai~ 0.0257 0.314 0.243 0.00686 0.0660 0.134 0.161
#> 3 Mauna Loa CO2     ETS_AAA  Test  2.83   3.34  2.83  0.687   0.687  1.71  1.87
#> 4 Mauna Loa CO2     ARIMA_1~ Test  2.60   3.06  2.60  0.631   0.631  1.57  1.72
#> # ... with 1 more variable: ACF1 <dbl>
```

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 Mauna Loa CO2     arima    2023 Jan  N(420, 0.1)  420. [419.8593, 420.6751]80
#> 2 Mauna Loa CO2     arima    2023 Feb N(421, 0.14)  421. [420.6271, 421.5950]80
#> 3 Mauna Loa CO2     arima    2023 Mrz N(422, 0.18)  422. [421.1692, 422.2439]80
#> 4 Mauna Loa CO2     ets      2023 Jan  N(420, 0.1)  420. [419.7793, 420.6008]80
#> 5 Mauna Loa CO2     ets      2023 Feb N(421, 0.14)  421. [420.4522, 421.4027]80
#> 6 Mauna Loa CO2     ets      2023 Mrz N(422, 0.17)  422. [421.0440, 422.1104]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 Mauna Loa CO2     arima    2052 Okt N(485, 106)  485. [471.7724, 498.1682]80
#> 2 Mauna Loa CO2     arima    2052 Nov N(487, 107)  487. [473.2884, 499.7706]80
#> 3 Mauna Loa CO2     arima    2052 Dez N(488, 107)  488. [474.6825, 501.2507]80
#> 4 Mauna Loa CO2     ets      2052 Okt N(485, 133)  485. [470.2634, 499.7980]80
#> 5 Mauna Loa CO2     ets      2052 Nov N(487, 134)  487. [471.7710, 501.4055]80
#> 6 Mauna Loa CO2     ets      2052 Dez N(488, 135)  488. [473.1340, 502.8684]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   2023     420.
#>  2 Mauna Loa CO2     arima   2023     421.
#>  3 Mauna Loa CO2     arima   2023     422.
#>  4 Mauna Loa CO2     arima   2024     423.
#>  5 Mauna Loa CO2     arima   2024     423.
#>  6 Mauna Loa CO2     arima   2024     424.
```

```
#>  7 Mauna Loa CO2     arima   2025     425.
#>  8 Mauna Loa CO2     arima   2025     426.
#>  9 Mauna Loa CO2     arima   2025     426.
#> 10 Mauna Loa CO2     arima   2026     427.
#> # ... 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   2023     418.
#>  2 Mauna Loa CO2     arima   2023     420.
#>  3 Mauna Loa CO2     arima   2023     421.
#>  4 Mauna Loa CO2     arima   2024     421.
#>  5 Mauna Loa CO2     arima   2024     422.
#>  6 Mauna Loa CO2     arima   2024     424.
#>  7 Mauna Loa CO2     arima   2025     423.
#>  8 Mauna Loa CO2     arima   2025     424.
#>  9 Mauna Loa CO2     arima   2025     426.
#> 10 Mauna Loa CO2     arima   2026     425.
#> # ... 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     23.4    0.800
#> 2 Mauna Loa CO2     ets       45.9    0.0318
```

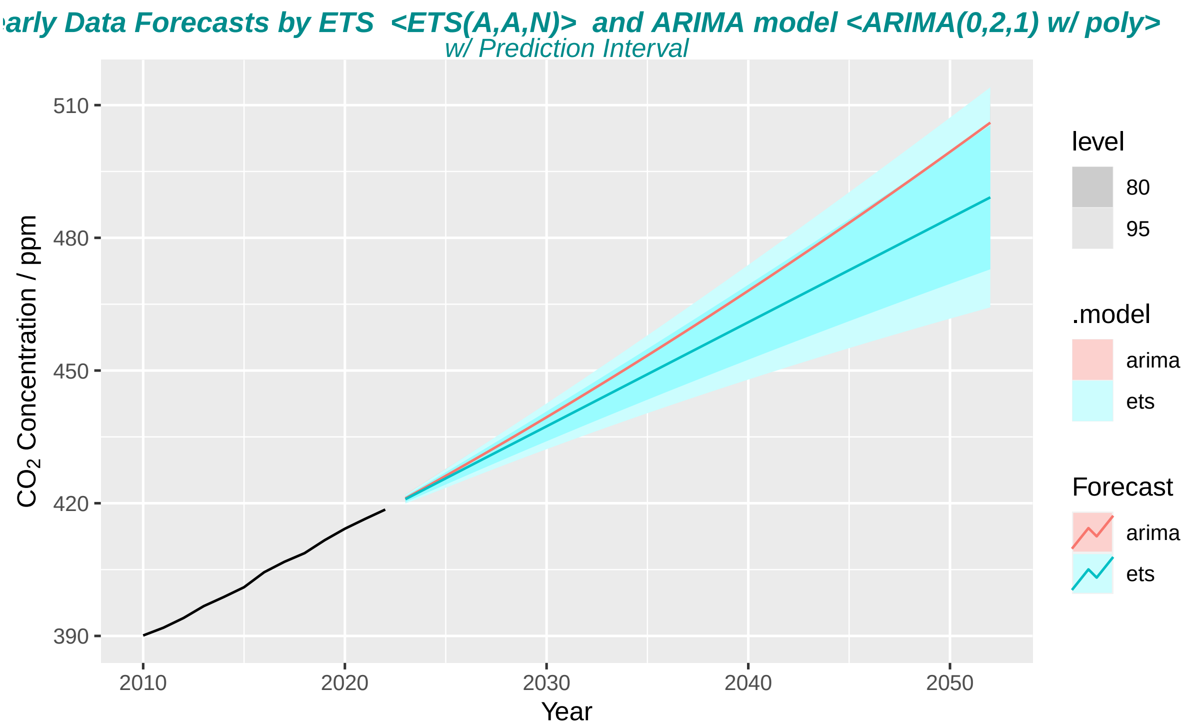
5 Yearly Data Forecasts with ARIMA and ETS

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.

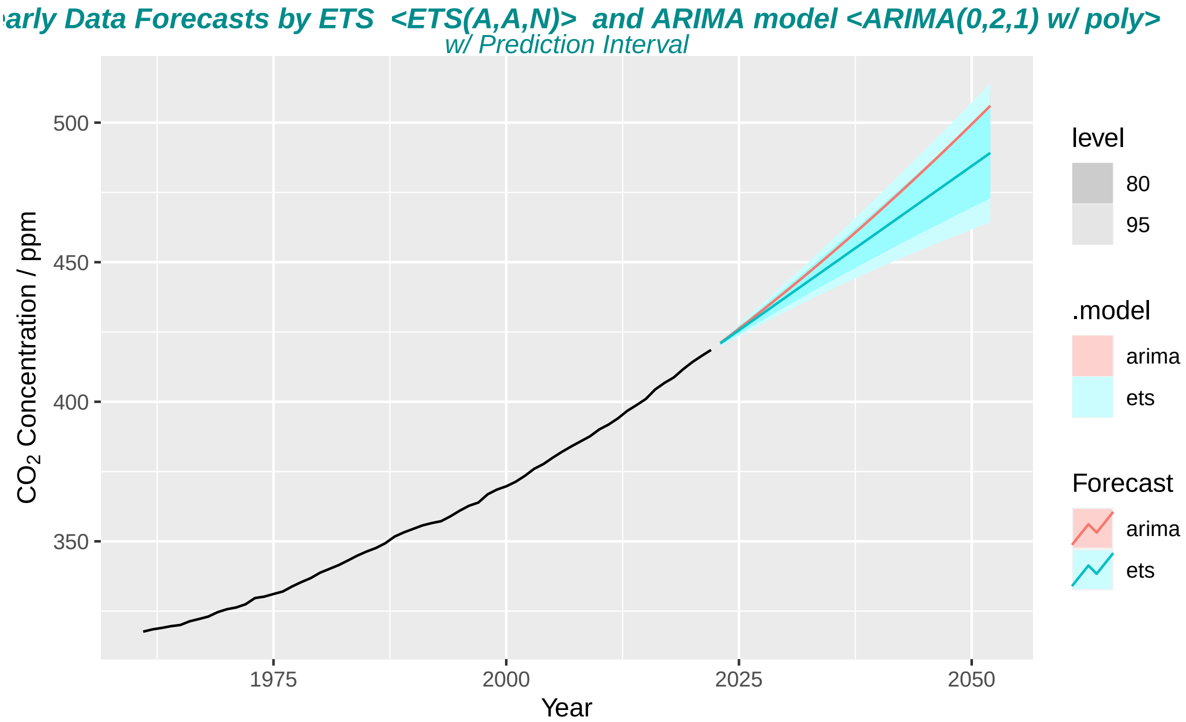
The ETS model $< ETS(A, A, N) >$ with seasonal term change "A" -> "N" is chosen. For ARIMA models the seasonal term (P,D,Q)m has to be taken out and an optimal ARIMA(p,1,q) with one differencing (d=1) is selected. However, for Mauna Loa two times differncing had to be selected $CO_2 <ARIMA(0,2,1) w/ 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 Mauna Loa CO2     arima   2023 N(421, 0.22)  421. [420.4719, 421.6794]80
#> 2 Mauna Loa CO2     arima   2024 N(424, 0.45)  424. [422.7527, 424.4773]80
#> 3 Mauna Loa CO2     arima   2025 N(426, 0.69)  426. [425.1159, 427.2487]80
#> 4 Mauna Loa CO2     ets     2023 N(421, 0.26)  421. [420.2620, 421.5725]80
#> 5 Mauna Loa CO2     ets     2024 N(423, 0.65)  423. [422.2389, 424.3018]80
#> 6 Mauna Loa CO2     ets     2025  N(426, 1.2)  426. [424.2291, 427.0176]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 Mauna Loa CO2     arima   2050 N(499, 9.6)  499. [495.4664, 503.3955]80
#> 2 Mauna Loa CO2     arima   2051  N(503, 10)  503. [498.6628, 506.7844]80
#> 3 Mauna Loa CO2     arima   2052  N(506, 11)  506. [501.8875, 510.2008]80
#> 4 Mauna Loa CO2     ets     2050 N(484, 134)  484. [469.5920, 499.3080]80
#> 5 Mauna Loa CO2     ets     2051 N(487, 147)  487. [471.2445, 502.3616]80
#> 6 Mauna Loa CO2     ets     2052 N(489, 161)  489. [472.8864, 505.4258]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     50.7  0.0104
#> 2 Mauna Loa CO2     ets       62.2  0.000489
```

6 Backup