20.6.2023

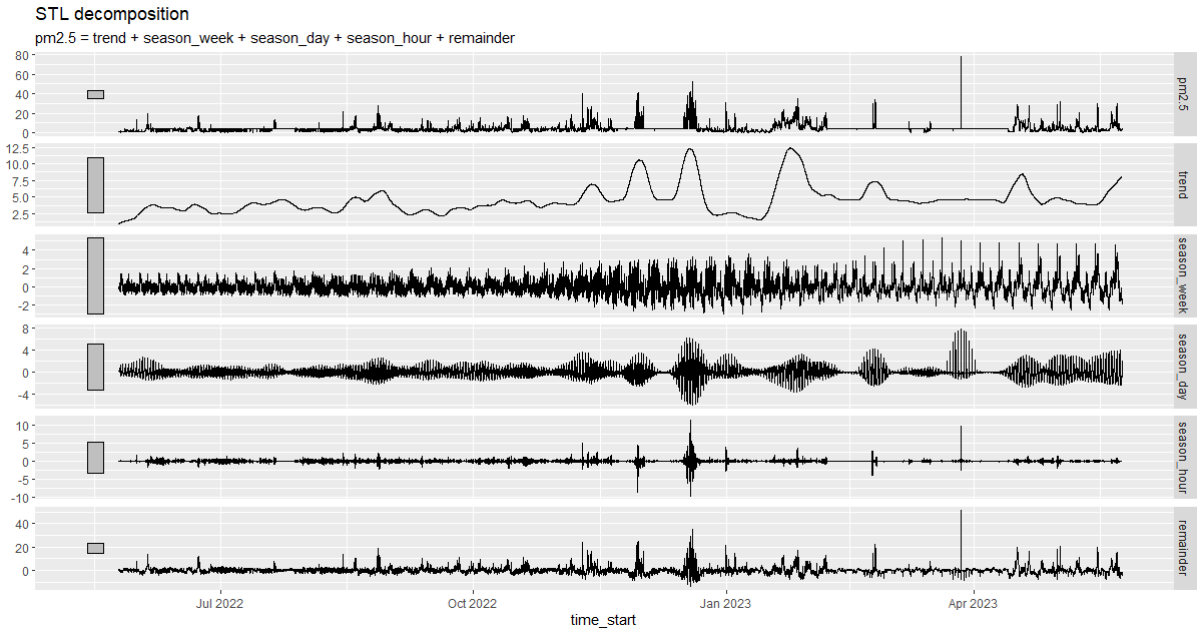
Gil Baram Geosensornetzwerke

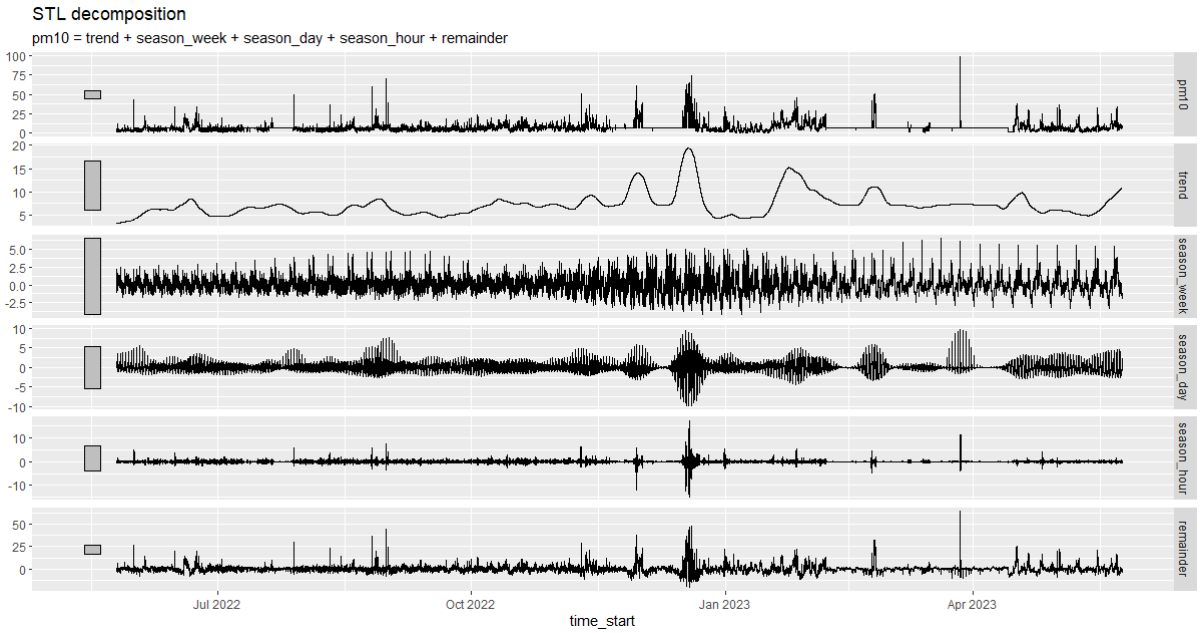
<u>Aufgabe 1: Evaluation der Daten</u>

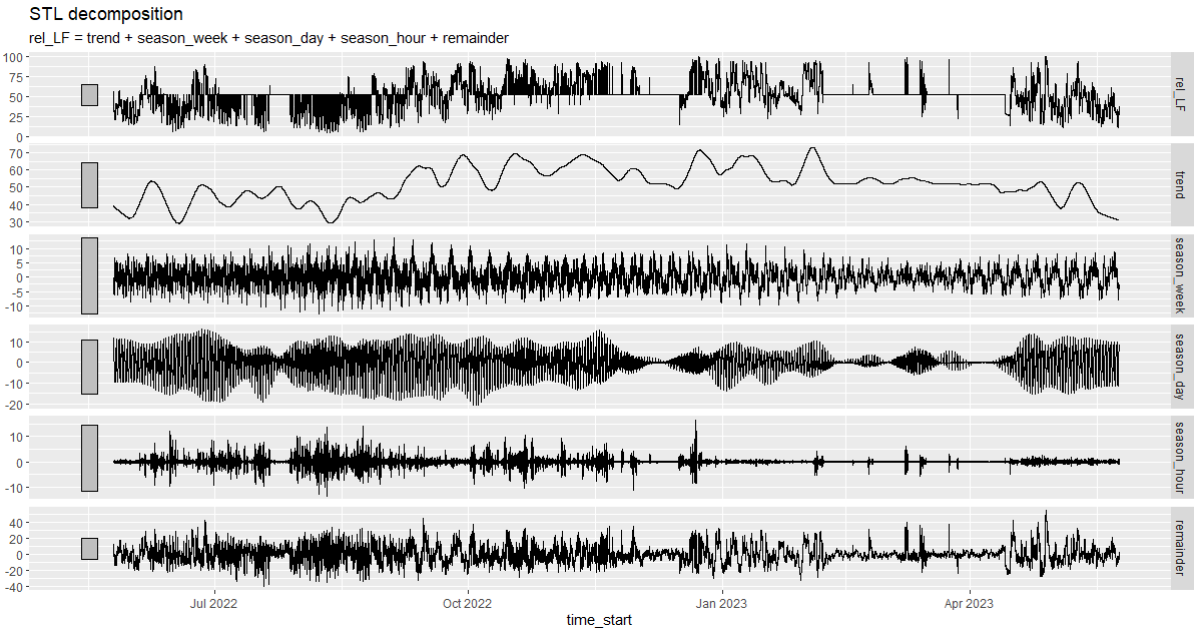
c)

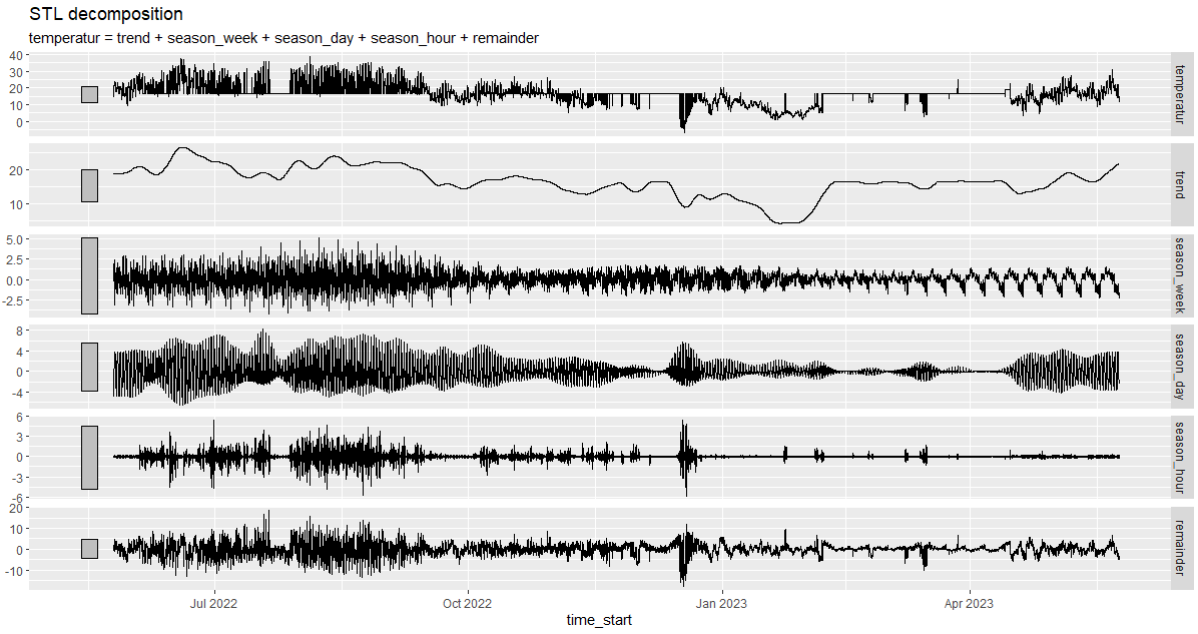
```
> summary(data)
  time_start            pm2.5              pm10             rel_LF          temperatur
 Length:31542       Min.   : 0.1925   Min.   : 0.225   Min.   : 4.625   Min.   :-6.75
 Class :character   1st Qu.: 1.6175   1st Qu.: 3.377   1st Qu.:35.133   1st Qu.:11.57
 Mode  :character   Median : 2.6913   Median : 5.185   Median :51.075   Median :16.45
                    Mean   : 4.5651   Mean   : 7.187   Mean   :51.779   Mean   :16.55
                    3rd Qu.: 5.0319   3rd Qu.: 8.467   3rd Qu.:68.146   3rd Qu.:21.37
                    Max.   :78.0300   Max.   :98.700   Max.   :99.900   Max.   :38.60
```

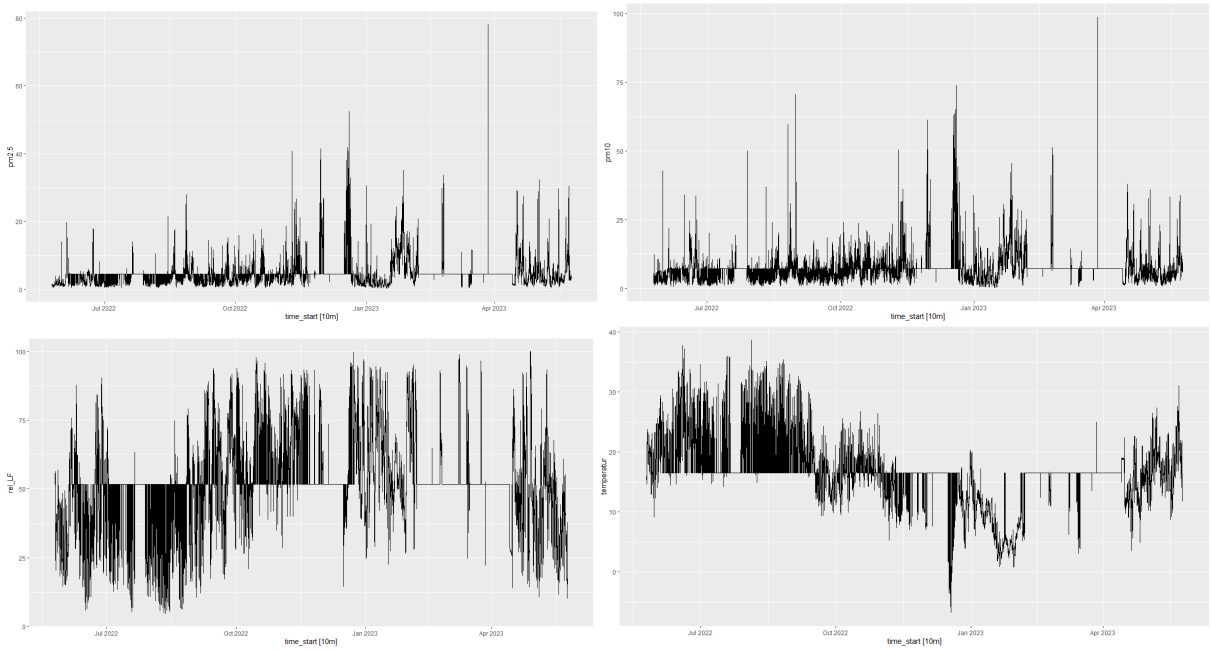
The data frame has 31542 rows, with no NA values. After applying gaps filling, it adds 21020 rows.

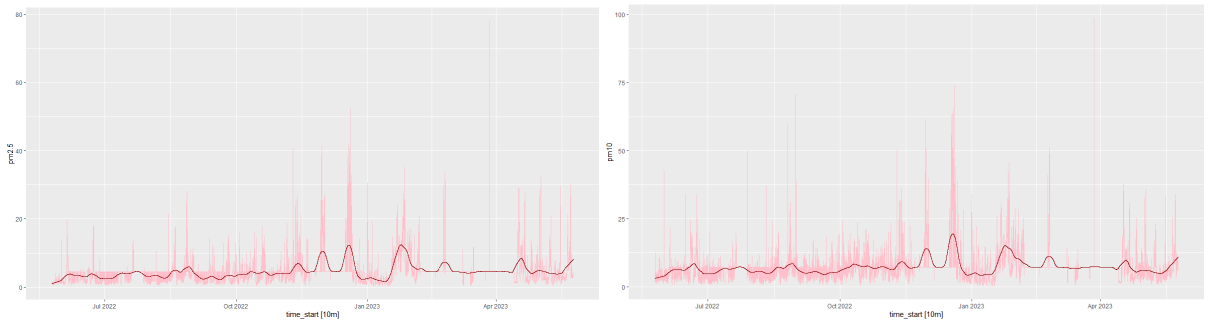




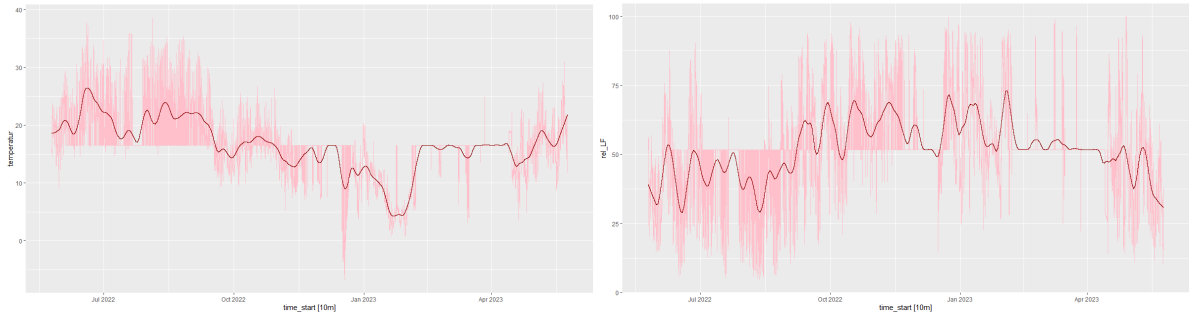

STL decomposition
rel_LF = trend + season_week + season_day + season_hour + remainder
rel_LF
trend
season_week
season_day
season_hour
remainder
Jul 2022
Oct 2022
Jan 2023
Apr 2023
time_start

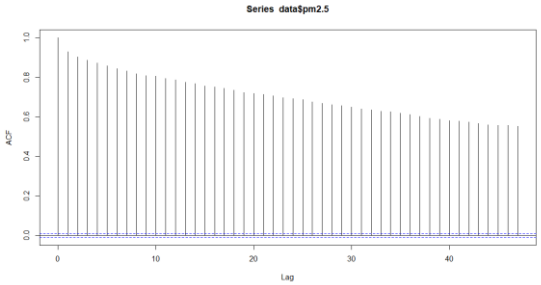

STL decomposition
temperatur = trend + season_week + season_day + season_hour + remainder
temperatur
trend
season_week
season_day
season_hour
remainder
Jul 2022
Oct 2022
Jan 2023
Apr 2023
time_start

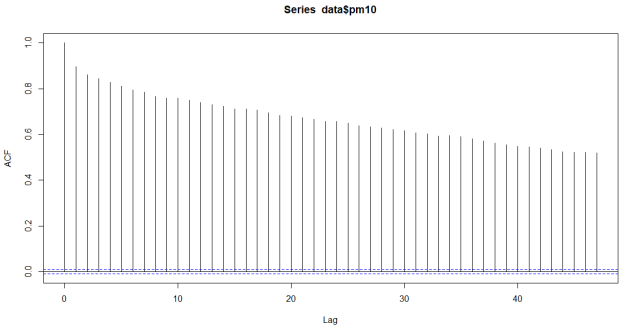


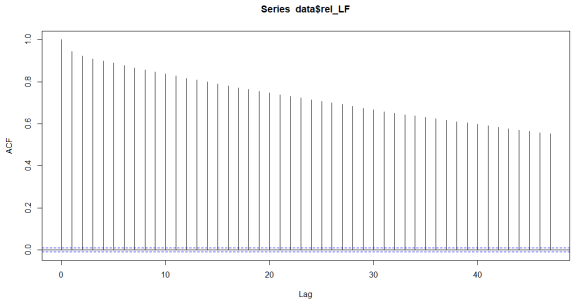


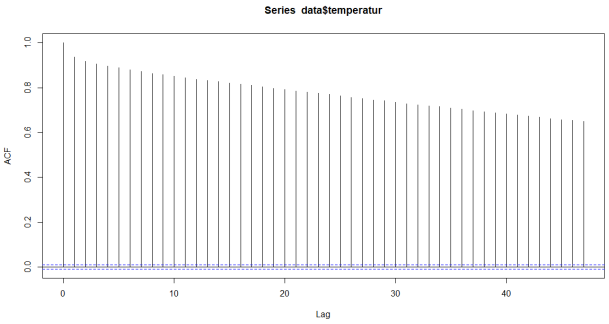
There is an outlier before Apr 2023. It reached a peak 80-100 µg/m³.



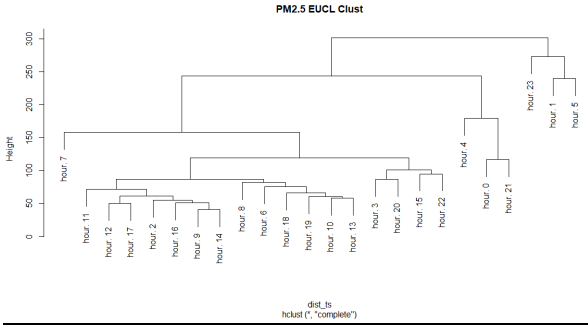

Series data$pm2.5
ACF
Lag

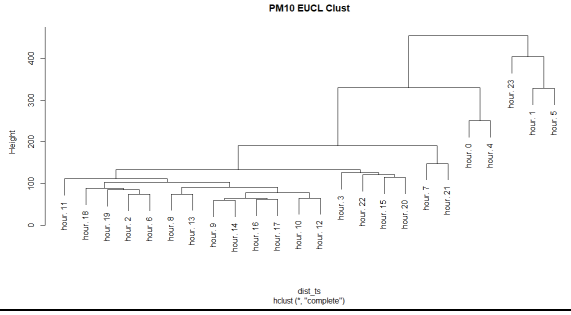

Series data$pm10
ACF
Lag

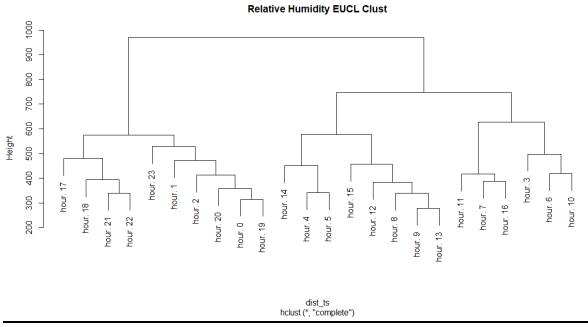

Series data$rel_LF
ACF
Lag

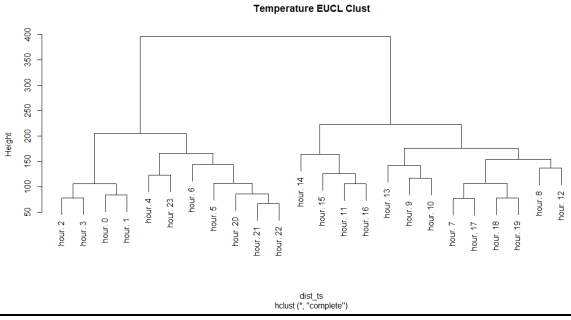

Series data$temperatur
ACF
Lag

Aufgabe 2: Zeitreihen Clustering



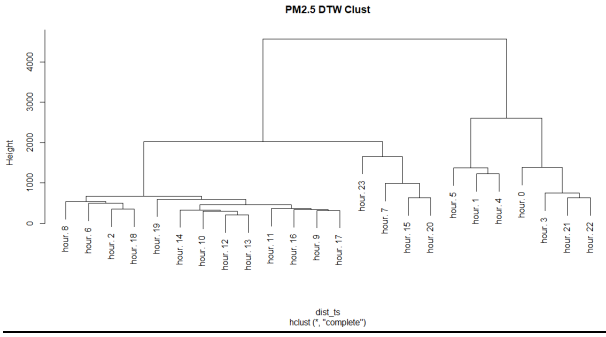


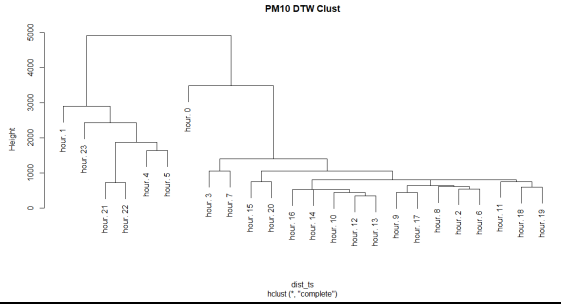


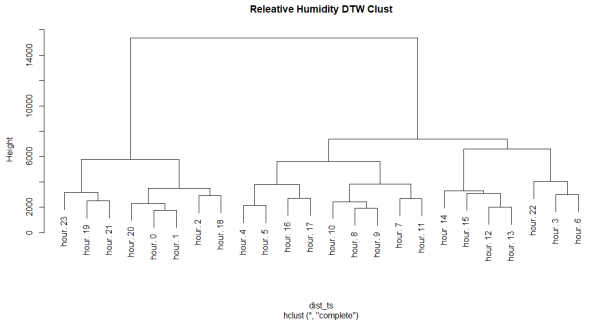


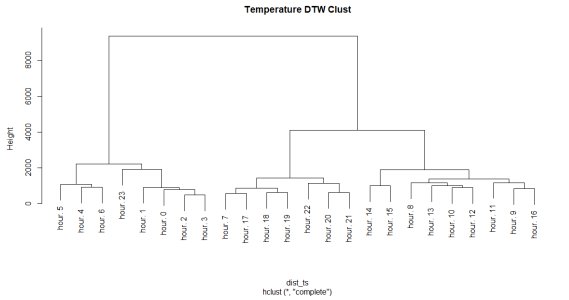
The Euclidean distance process determines the proximity between observations by drawing a straight line between pairs of observations. Therefore, this process measures the distance between observations by looking at the length of this line between observations.

The clusters of PM2.5 and PM10 are identical. The measures during the night (hours = 1,5,23) are at same level, at this is the largest cluster.









In hierarchical clustering, the algorithm builds clusters by measuring the dissimilarities between data.

Aufgabe 3: Zeitreihen Forecasting

c) I calculated MAE for three different Forecasts for all the 4 features.

```
> fit_CV<- data_stretch |> model(MEAN(pm2.5))
> mae_display(fit_cv)
[1] 2.231458
[1] 2.288092
[1] 2.156506
> fit_CV<- data_stretch |> model(MEAN(pm10))
> mae_display(fit_cv)
[1] 2.781453
[1] 2.848204
[1] 2.72481
> fit_CV<- data_stretch |> model(MEAN(rel_LF))
> mae_display(fit_cv)
[1] 12.22813
[1] 12.10388
[1] 11.82352
> fit_CV<- data_stretch |> model(MEAN(temperatur))
> mae_display(fit_cv)
[1] 4.205818
[1] 4.133279
[1] 4.153775
```