CBS COPENHAGEN BUSINESS SCHOOL
HANDELSHØJSKOLEN

Prediction of Temperature Development in Denmark

Course name: Predictive Analytics
Programme: MSc in Business Administration and Data Science
Examination type: Home Assignment – Written Product
Student name: Faye Sabine Hahn (149875)
Supervisor: Andrea Tafuro

Date of Submission: 22.08.2022

Characters: 19.144
Pages: 10

Inhaltsverzeichnis

1 Introduction

The average global temperatures have risen considerably since the start of the industrial revolution and the last decade was the warmest decade that was ever recorded (European Parliament, 2022). The first effects of this on Europe can already be seen in various forms like forest fires and heat waves. And even if Denmark is not the first country to be affected by this development due to its geographical position, it will probably sooner or later experience similar issues. Therefore, it is important to have the right tools at hand to predict the temperature development in order to prepare for the potential challenges that are associated with this.

This paper therefore aims to find a suitable model to predict the further development of the temperature in Denmark. This is done by using data on the monthly average temperature in Denmark covering the last 10 years (2012-2021) and fitting different models from two complementary model classes (ARIMA and ETS) on it. These models will then be compared to each other to ultimately find a model that will be used to forecast the temperature development for 2022 and 2023.

The study showed that among the considered models an ETS(A,N,A) was the most suitable to forecast temperature development in Denmark. According to the forecast, a rapid increase of the temperature is not to be expected for the current or the next year. However, there is still potential for improving the fit of the model in order to produce more precise and reliable predictions in the future.

2 Data Description and Preliminary Analysis

This chapter is about gaining a deep understanding of the available data while also preparing for the model creation in Chapter 3. Therefore, the original data is briefly discussed in the first part of this chapter which is followed by a deeper analysis of the different components and features in the data.

2.1 Data Description

The dataset used for this project was obtained from the Climate Change Knowledge Portal (CCKP) which is part of the World Bank Group. The Timeseries covers monthly data for the period of 1901 to 2021, capturing the development of the average temperature in Denmark in degree Celsius. For the sake of this study, the dataset was reduced to a sample of the 120 most recent data points, covering only the period from January 2012 until December 2021. Still, even for this shortened period, the dataset allows capturing recent patterns which is relevant for predicting temperature development for the next couple of years. Thereby, the values within this sample range between -1.04°C and 19.24°C with a mean value of 9.261°C.

2.2 Preliminary Analysis

Before creating the models to generate forecasts, the data was examined more closely in order to capture all relevant characteristics and features within it. As a first step, the data was plotted which can be seen in Figure 1.

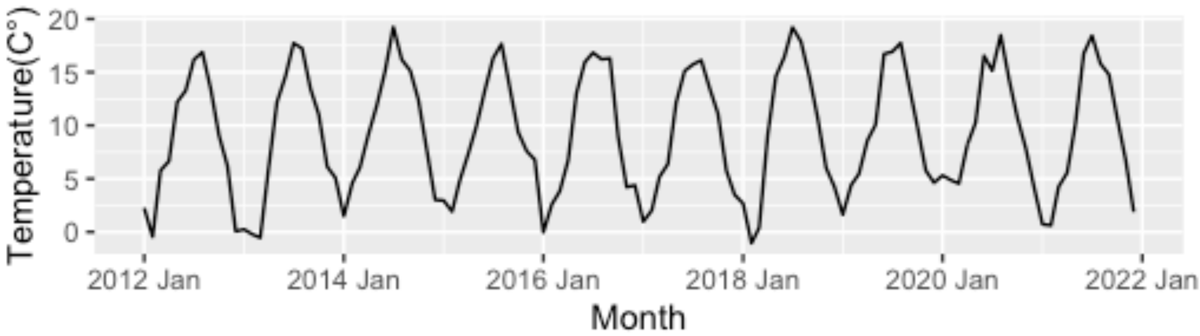


Figure 1: Temperature development in Denmark 2012-2021

The plot showed strong seasonality which further made it difficult to spot any other features in the data like a trend. Furthermore, the plotted data did not seem to include any structural breaks.

The next step was a mathematical transformation of the data using the Box-Cox method. This allowed for easier forecasting by bringing the data distribution closer to a normal distribution and in this case align the seasonal variation across the whole dataset. The Guerrero method which was used to find a suitable lambda value for the transformation ultimately produced a value of 1.720236 which indicates that a power transformation was applied.

After the transformation, the data was ready to be decomposed. This was done to allow for a closer examination of the different components in the data. Because of the strong presence of seasonality in the first plot, an STL decomposition was applied here which is an acronym for “Seasonal and Trend decomposition using Loess”. The result from the decomposition is visualized in Figure 2.

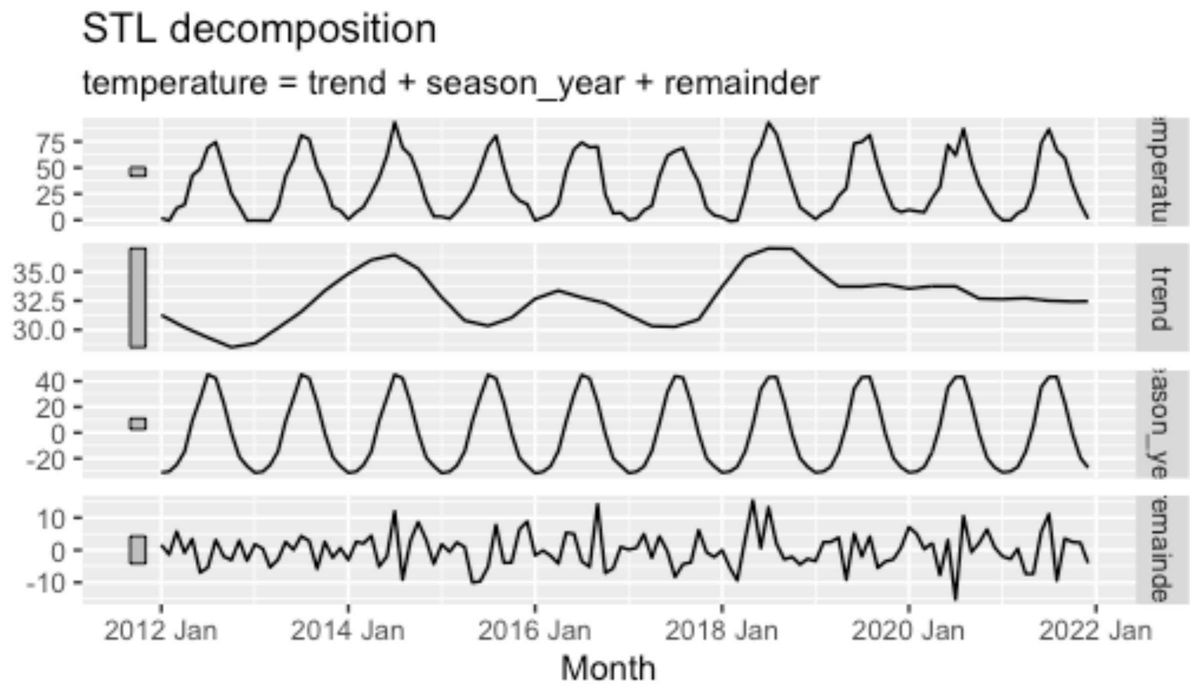


Figure 2: STL decomposition

The first thing that could be observed from this is how the scale of the original data got stretched from a range between approximately -5 and 20 degrees to a new range up to approximately 100 degrees.

The data did not seem to contain a clear trend for the sample period. Upon closer examination, it could be seen that it rather fluctuated around a constant (mean) value of 32.73 degrees with an amplitude of maximum 4.2 degrees[1]. Told from its course, this could potentially have been some type of cyclic behavior, while the small magnitude of its fluctuations spoke against this theory. Due to the limited horizon of this paper, this was not investigated further at this point, but taken into account for the reflection on the models in Chapter 5.

As anticipated, the seasonality seemed to be the main component in the data. Even though, the STL method would allow for a varying seasonality over time, it seemed to be constant over time which was probably partly due to the applied Box-Cox transformation.

Lastly, the error looks okay, but still seems to contain some level of heteroskedasticity after the transformation. 2012 until the beginning of 2014 showed relatively small errors at a rather constant level while it became stronger with higher variation afterwards.

Next, the data was checked for stationarity. This is important since unspotted stationarity can cause forecasting models to perform worse and some models like ARIMA even rely on the data being stationary to work properly. To identify whether the data is stationary, a combination of the Augmented Dickey-Fuller (ADF) test and the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test was used. The ADF test thereby served for a first assessment of whether the data contained a unit root (i.e. was non-

[1] Please note that the denoted temperature values both refer to temperatures on the transformed scale and not on the original one.

stationary) (H_0), while the KPSS test was mainly used to confirm the results of the ADF test by indicating whether the data is stationary (H_0) or not (H_1). As no consistent trend could be spotted in the data, an ADF test of the type "drift" was applied first. The test produced a t-value which was smaller than the critical value on a 1% significance level, meaning H_0 could be rejected, and the data was potentially stationary. A repetition of the ADF test – this time with type "None" – produced a similar outcome. Finally, the null hypothesis of the KPSS test could not be rejected which confirmed the results from the ADF tests of the data being stationary already, without the need of being differenced first. The full results from the executed tests can be found in Appendix A.

As a last step of the analysis, the data was tested on structural breaks. Even though a visual analysis of the plot in Figure 1 did not reveal clear signs of any breaks in the data, a Quandt Likelihood Ratio (QLR) test was executed in addition to confirm this Hypothesis from a statistical standpoint. The test could not identify any endogenous breaks in the data which rendered any truncation of the time series prior the forecast generation obsolete. A plot, visualizing the test result can be found in Appendix B.

3 Methodology

This section is devoted to the definition of the models that were then used for the forecast generation on the temperature development in Denmark. Two high-level model classes were used here which is the ETS model on the one hand and an ARIMA model on the other hand. Both models were chosen mainly due to their wide use and their ability to model seasonal data. Additionally, both models follow complementary approaches which makes a comparison of their forecasting performance even more interesting. In both cases, at least one self-created model was set up and compared to one automatically generated benchmark model. The best model from each class was then evaluated again, this time based on its forecasting quality and ultimately, the overall best model was selected and used for the final forecast of the temperature development in 2022 and 2023.

3.1 Train test split

Before the models were created, the data was split into a train and a test set first. This allowed to measure the predictive accuracy of the models on held out i.e. unseen data. For this study, a train to test ratio of 80:20 was used which ultimately produced a train set covering the period from January 2012 to December 2019, while the test set contained the data of the remaining two years from January 2020 to December 2021.

3.2 ETS Model Creation

Exponential Smoothing was proposed already in the 1950s and has been widely used in industry ever since. The ETS model combines Error, Trend and Seasonality in a smoothing calculation in an either additive or multiplicative manner. If a component is not present i.e. relevant for the data modeling, it can also be omitted from the model (Hyndman & Athanasopoulos, 2021).

For the self-created model, the first component to be discussed was the error term. Because of the Box-Cox transformation that was applied to the data in the beginning, an additive error term was included in the model. Next, the trend term was omitted which was based on the findings from the decomposition in Chapter 2.2. Lastly, an additive seasonal term was included. This was also derived from the result of the decomposition which did not indicate any variations in the amplitude of the season component over time. The final self-created model (i.e. *ets_guess*) therefore was an ETS(A,N,A). A comparison of the self-created and the auto-generated model can be found in Table 1.

	ets_guess	**ets_auto**
	ETS(A,N,A)	ETS(A,N,A)
AIC	826.5716	826.5716
AICc	832.4975	832.4975
BIC	865.1922	865.1922

Table 1: Comparison of ETS models

The first thing that stood out was that the auto-generated model (i.e. *ets_auto*) had the same architecture as the self-created model which rendered a comparison of the two models obsolete. This did not have to be a negative sign but was taken into consideration during the examination of properties in the next step.

α	γ	l_0	s_{0-12}
0.131	0.0001	30.78608	Between -30.7269 and 42.90958

Table 2: Coefficients of the ETS model

Table 2 shows an overview of the estimated smoothing parameters as well as the initial states for the model. Both, alpha and gamma value indicated a high inertia in the model, meaning older observations had a long lasting impact on future predictions. The extremely small value for the seasonality once again highlighted the predominance of the seasonal component in the model.

Another source for information about the suitability of the model was the residuals plot which can be found in Figure 3.

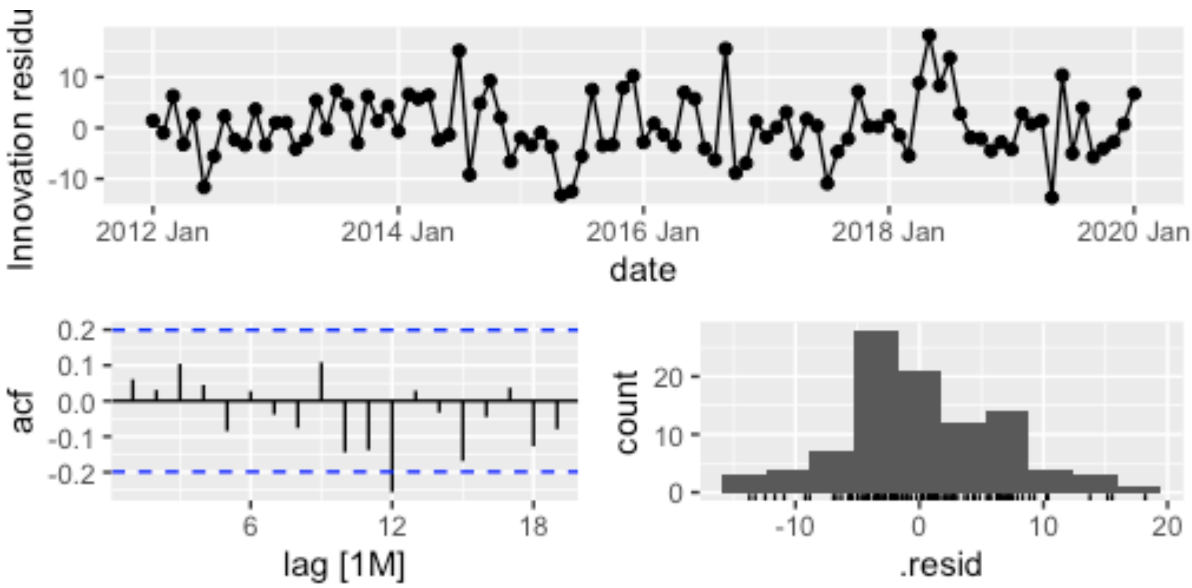


Figure 3: Residuals of the ETS model

Overall, the plot did not reveal a good picture with respect to the residuals. Although, the ACF showed only one significant spike at lag 12, the residuals showed some level of heteroskedasticity over time. Additionally, the residuals' distribution did not seem to correspond to a normal distribution. The Ljung-Box test confirmed this impression by rejecting the null hypothesis of independent residuals on a 5% level. This implied that not all variation in the data could be captured by the chosen ETS model.

In order to improve this, some additional experimentation on the ETS model creation was done, but ultimately, the selected ETS(A,N,A) model still remained the best performing model on the train set while additionally being in accordance with the findings from the Analysis part.

3.3 ARIMA Model Creation

The second class of models used in this paper is ARIMA which is an abbreviation for "autoregressive integrated moving averages". In contrast to ETS models, ARIMA models aim to describe the autocorrelations in the data in order to generate predictions. This is done by using three main components: an autoregressive component (AR), an integration (I) and a moving average (MA). Thereby, all three components can be included on an either non-seasonal or seasonal level (Box et al., 2016).

In order to determine the different components of the ARIMA model, the ACF and PACF of the stationary timeseries needed to be examined more closely. The results of the analysis in Chapter 2.2 have shown that the undifferenced timeseries was already stationary and would therefore require no further differencing. But on the other hand, the model needed to account for the seasonality in the data by including a seasonal component. Therefore, as a rule of thumb, seasonal differencing with a lag of twelve for monthly data was applied. The resulting plots showing the ACF and PACF for the seasonally differenced data are shown in Figure 5.

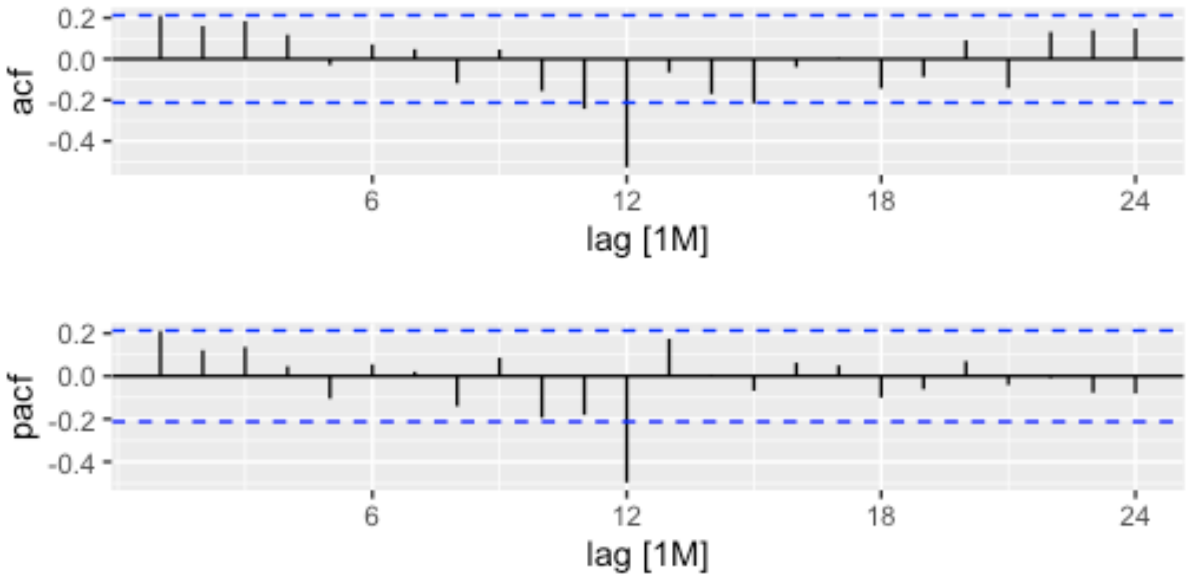


Figure 4: ACF and PACF plot of the seasonally differenced data

The most striking feature in both ACF and PACF was the significant spike at lag 12. Its orientation indicated an overdifferencing of the series on a seasonal level, so we added a seasonal MA(1) component to mitigate this effect. Furthermore, it could be seen that the spike at lag 1 was slightly significant in both plots. As neither an AR(0) nor an AR(1) component could be ruled out, both options were taken into account for the model comparison. The ACF also featured a significant spike at lag 11, but as this did not seem to be part of any clear pattern, no additional measures were taken to account for this. Finally, a drift term was included in both self-created models which is due to both, the results from the ADF test as well as the observations about the trend from the decomposition. In summary, two models were created – one being an ARIMA(0,0,0)(0,1,1) with drift (named *arima_guess1*) and the other being an ARIMA(1,0,0)(0,1,1) with drift (named *arima_guess2*). The comparison of the two self-created models to the auto-generated model (i.e. *arima_auto*) can be found in Table 4 below.

	arima_guess1 ARIMA(0,0,0)(0,1,1)[12] w/ drift	**arima_guess2** ARIMA(1,0,0)(0,1,1)[12] w/ drift	**arima_auto** ARIMA(1,0,0)(0,1,2)[12]
AIC	589.11	589.29	590.18
AICc	589.41	589.79	590.68
BIC	596.44	599.06	599.95

Table 4: Comparison of ARIMA models

The auto-generated ARIMA model also included a non-seasonal AR(1) component, but other than the two self-created models, it featured a seasonal MA(2) rather than an MA(1) component and did not include a drift term. Overall, the table revealed that all three models achieved similar scores across the three Information Criteria that were used for the comparison. The best values in terms of all three Criteria were achieved by the first self-created model. A look at the significance levels for each of the coefficients furthermore showed that both, *arima_guess2* and *arima_auto*, contained components that could be omitted, while in *arima_guess1* all components were significant. Altogether, the ARIMA(0,0,0)(0,1,1) with drift seemed to be the most suitable model.

Again, a residuals plot was used in order to assess the ability to account for variation in the data. The plot can be found in Figure 5.

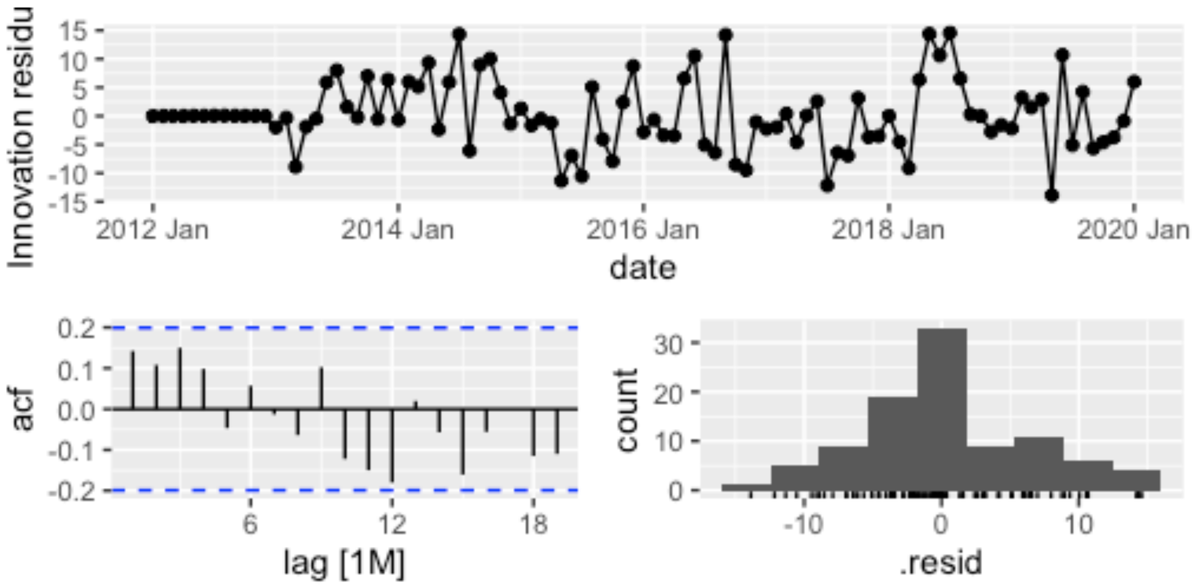


Figure 5: Residuals of the ARIMA model

Judging from the plots, the residuals for the ARIMA model looked slightly better than for the ETS model. Not only were there no more significant lags in the ACF plot, but the distribution got also closer to what looked like a normal distribution, even though it still showed some level of skewness. Still, there was again some heteroskedasticity in the residuals over time and even though the result of the Ljung-Box test got slightly better, it failed to reject the null hypothesis of independent lags on a 10% level. Altogether, the chosen ARIMA model seemed to be capable to capture some of the variance, but also misses out on some of it.

4 Results

Finally, we were ready to compare the models on their ability to generate forecasts on temperature development. Therefore, the models were retrained with the untransformed data and then used to generate forecasts for the test period of 2 years. The result of this is visualized in Figure 6.

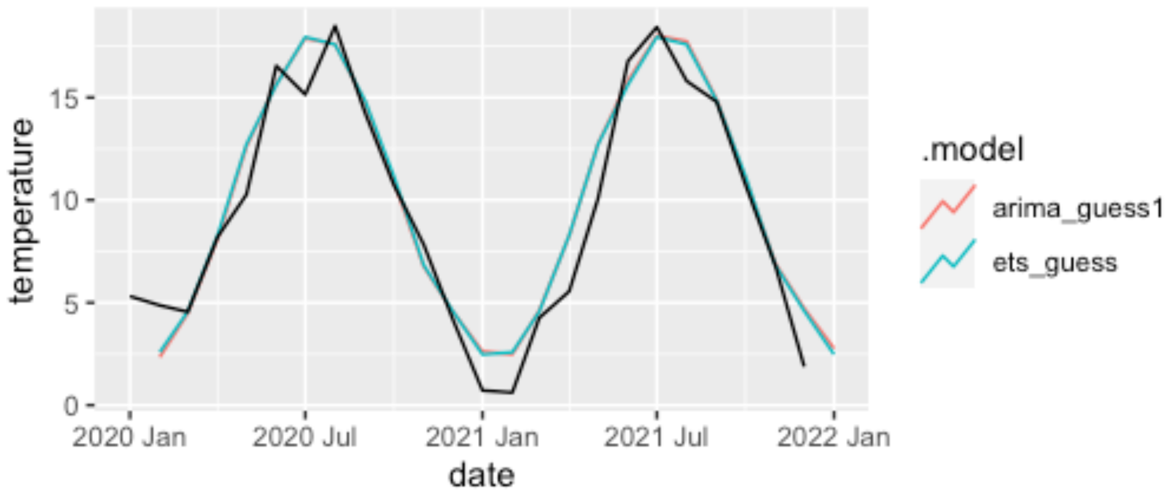


Figure 6: Forecasts of both models on the test set

Both models produced very similar forecasts for the test period. It could be seen that they started at a lower point than the actual plot and showed a slightly smaller amplitude in the swings. Especially, in the down swing at the beginning of 2021, a clear difference between the forecasted and the actual temperature could be observed. Overall, it was hard to tell which model performed better on the forecast from a visual perspective. The individual plots revealed a similar picture, but in direct comparison, it could be observed that the ARIMA model produced a slightly smaller confidence interval for the forecast (see Appendix C).

To get a better idea which model had a higher ability to account for variation in the test set, Table 4 summarizes the main error metrics.

	RMSE	MAE	MAPE	MASE
ARIMA	3.46	3.02	73.2	1.69
ETS	3.44	3.00	72.8	1.68

Table4: Forecast errors for both models on the test test

The comparison of the forecast errors for the two presented models confirmed the observations from the visual analysis about the high similarity of the models. Still, the ETS model ultimately achieved lower errors across all four error types, rendering it the better model for the sake of temperature forecasting.

Finally, Figure 7 presents the plot of the final forecast for the years 2022 to 2023, generated with the ETS(A,N,A) model which seemed to capture most patterns in the data on a decent level, leading to a fairly realistic looking forecast that still included some level of uncertainty. The resulting prediction implied that the temperature might stay on a similar level as before, but also considered possible deflections, both up or down.

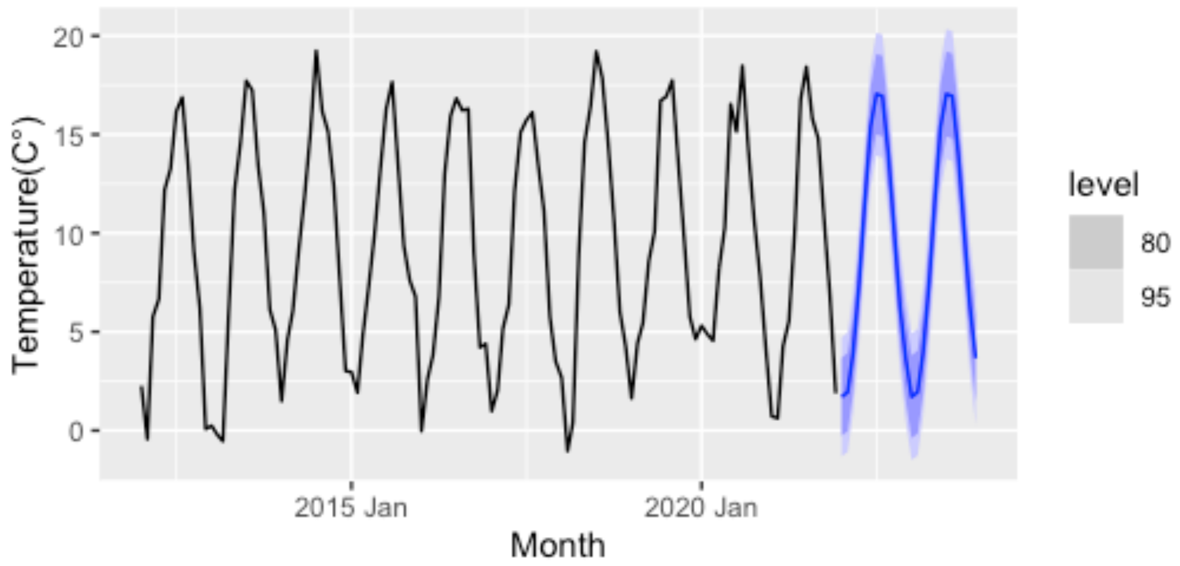


Figure 3: Final forecast for the temperature development in Denmark 2022-2023

5 Conclusion

The aim of this paper was to find a suitable model to predict the temperature development in Denmark. Therefore, a total of five[2] models from the classes of ETS and ARIMA was fitted to the data and evaluated based on their ability to capture the patterns in the dataset. The best performing model out of this selection was an ETS(A,N,A) model, although its forecasts still featured relatively large confidence intervals i.e. an increased level of uncertainty and was also considerably far from accounting for every type of variation in the data.

Therefore, future research in this field could take up on the results of this paper to explore other models and maybe even variables that could facilitate producing more reliable predictions. Furthermore, the number of data points taken into consideration should be increased to also take long term developments like cycles into account. The amount of data could also be increased by using higher frequented data to predict the development on a higher granularity.

Lastly, it is to be noted that this project only looked at average values which might blur the existence of some stronger outliers that could indicate days of extreme heat. Thus, it could also be interesting to have a look at maximum temperatures instead of average temperatures.

Nonetheless, looking at the predictions for the current and the upcoming year, Denmark can probably still enjoy the one or other summer break with rather pleasant Scandinavian temperatures.

[2] Four models, if we disregard the duplicated ETS model

References

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2016). *Time series analysis: Forecasting and control* (Fifth edition). John Wiley & Sons, Inc.

European Parliament. (2022, June 17). *EU responses to climate change*. https://www.europarl.europa.eu/news/en/headlines/society/20180703STO07129/eu-responses-to-climate-change. Last accessed: 21-08-2022.

Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and practice* (3rd edition). Otexts, online, open-access textbook.

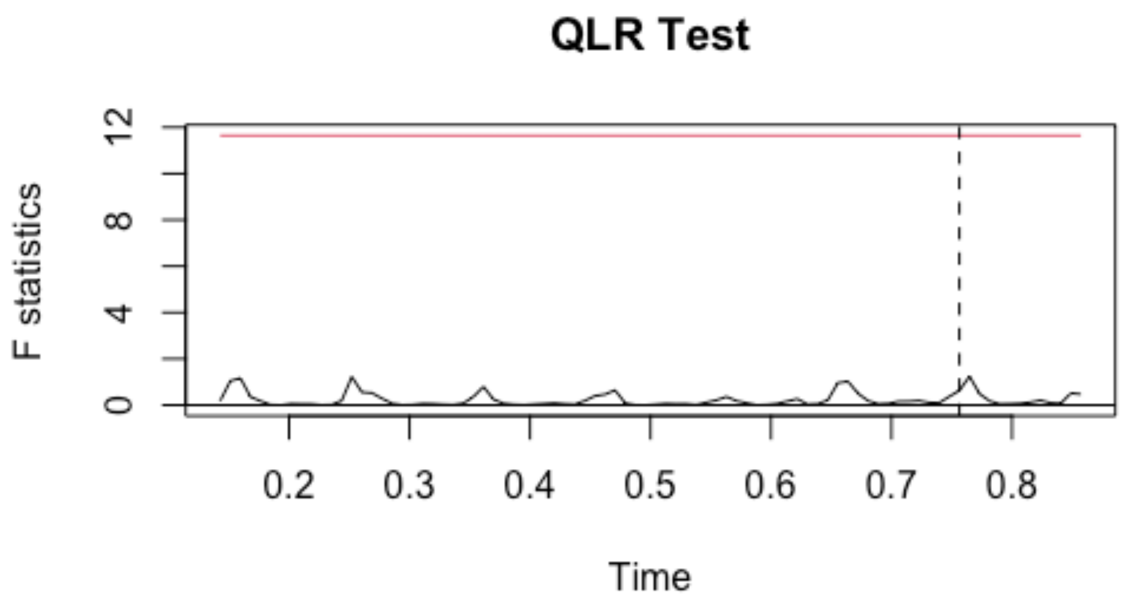
Appendix

Appendix A: ADF and KPSS test results

```
###############################################
# Augmented Dickey-Fuller Test Unit Root Test #
###############################################

Test regression drift


Call:
lm(formula = z.diff ~ z.lag.1 + 1 + z.diff.lag)

Residuals:
    Min      1Q  Median      3Q     Max 
-34.434  -8.279  -1.091   4.562  40.876 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 10.33810    1.92642   5.366 4.22e-07 ***
z.lag.1     -0.31274    0.04504  -6.943 2.43e-10 ***
z.diff.lag   0.62767    0.07260   8.645 3.66e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 13.2 on 115 degrees of freedom
Multiple R-squared:  0.4529,	Adjusted R-squared:  0.4434 
F-statistic: 47.61 on 2 and 115 DF,  p-value: 8.651e-16


Value of test-statistic is: -6.9429 24.1022 

Critical values for test statistics: 
      1pct  5pct 10pct
tau2 -3.46 -2.88 -2.57
phi1  6.52  4.63  3.81

###############################################
# Augmented Dickey-Fuller Test Unit Root Test #
###############################################

Test regression none


Call:
lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)

Residuals:
    Min      1Q  Median      3Q     Max 
-37.199  -2.335   3.370  10.927  42.686 

Coefficients:
           Estimate Std. Error t value Pr(>|t|)    
z.lag.1    -0.12520    0.03165  -3.957 0.000131 ***
z.diff.lag  0.53712    0.07862   6.832 4.11e-10 ***
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 14.7 on 116 degrees of freedom
Multiple R-squared:  0.3159,	Adjusted R-squared:  0.3041
F-statistic: 26.79 on 2 and 116 DF,  p-value: 2.725e-10


Value of test-statistic is: -3.9565

Critical values for test statistics:
      1pct  5pct 10pct
tau1 -2.58 -1.95 -1.62

#######################
# KPSS Unit Root Test #
#######################

Test is of type: mu with 4 lags.

Value of test-statistic is: 0.0256

Critical value for a significance level of:
                10pct  5pct 2.5pct  1pct
critical values 0.347 0.463  0.574 0.739
```

Appendix B: QLR test plot



Appendix C: Individual forecasts on the test set